Research Report

Measures of Agreement Versus Measures of Prediction Accuracy

ETS RR–19-20

Shelby J. Haberman

December 2019

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Measures of Agreement Versus Measures of Prediction Accuracy

Shelby J. Haberman

Consultant, Jerusalem, Israel

Measures of agreement are compared to measures of prediction accuracy within a general context. Differences in appropriate use are emphasized, and approaches are examined for both numerical and nominal variables. General estimation methods are developed, and their large-sample properties are compared.

Keywords Kappa; lambda; correlation; proportional reduction of error; coefficient of determination

doi:10.1002/ets2.12258

Measures of agreement and measures of prediction accuracy are both commonly used. They serve different purposes, so it is important to understand their proper use. It is also important to understand the options for measuring agreement and the options for measuring prediction. This report provides a series of recommendations for their use within educational assessment and the basis for these recommendations. The report is not comprehensive to the extent that reasonable measures are omitted, including measures that the author has employed in the past. In addition, the methodology considered here is generally not new, although some new results were needed to address some apparent omissions in the literature detected in preparation of the report. In the section Measures of Agreement, measurement of agreement is examined, while measurement of prediction accuracy is examined in the section Measures of Prediction Accuracy. Estimation is considered in the section Estimation of Measures in terms of estimates, large-sample approximations for distributions, and approximate confidence intervals. Some concluding remarks are provided in the section Concluding Remarks. Although many of the arguments required for a general discussion are relatively technical, a number of basic points can be made prior to consideration of details.

Measures of agreement are symmetrical assessments of whether two or more measurements can be regarded as interchangeable. Within educational testing, these measures have often been applied to rating constructed responses, and they have often involved variables that can assume only a finite number of values. Common practice has often involved probabilities of exact agreement or exact or adjacent agreement (Goodman & Kruskal, 1954) or measures from the kappa family (Cicchetti & Allison, 1971; Cohen, 1960, 1968; Conger, 1980; Fleiss & Cohen, 1973; Fleiss, Cohen, & Everitt, 1969); however, related but somewhat different measures based on a different application of proportional reduction of error (Costner, 1965; Goodman & Kruskal, 1954) merit attention. For discrete or continuous variables, agreement can be assessed by use of intraclass correlation (Fisher, 1934, pp. 199–203). Whatever agreement measures are used, they should be applied to similar kinds of measurements, such as scores from different human raters randomly selected from a pool or similar thermometers, for the measures should be plausibly interchangeable.

As is the case for all common measures, except for the probability of exact agreement or the probability of exact or adjacent agreement, good practice generally involves proportional reduction in error (Costner, 1965), so that a measure of the discrepancy of the measurements, such as mean absolute difference, is compared to a baseline measure that reflects the extent to which the individual measurements vary. In the kappa family, the baseline measure is the measure of discrepancy that would result if the variables under study were independently distributed. In the lambda family for agreement based on Goodman and Kruskal (1954), the baseline is the average discrepancy of each variable under study compared to the average discrepancy of each variable compared to an optimally selected constant. The lambda measure for agreement is smaller than the corresponding kappa measure, except in a few special cases in which the measures both attain their maximum value of 1.

Corresponding author: S. J. Haberman, E-mail: haberman.statistics@gmail.com

The following issue leads to proportional reduction of error as a needed principle. If all raters assign the same score to all essays being rated, then they agree with each other well simply because no variation in scores exists. As an example of this basic result leading to a problem in practice, exact agreement can be increased by grouping of variable values, a practice that may be questionable. For example, consider two methods to ascertain native language. The exact agreement of these methods is increased by grouping languages together so that, for example, French, Spanish, and Italian are combined into a category for Romance languages; however, no actual improvement in consistency has been observed.

Both the kappa measures and the lambda measures for agreement are defensible measures given their use of proportional reduction of error; however, the kappa measures are much more familiar to psychometricians than the lambda measures for agreement. Within kappa measures, a distinction should be made between agreement for nominal variables, such as native language, and agreement for numerical variables, such as holistic score. For nominal variables, the essential issue is whether the variables are the same, so that exact agreement and the original kappa measure (Cohen, 1960) are appropriate. In the case of numerical variables, a reasonable measure of discrepancy, such as mean absolute difference (Cicchetti & Allison, 1971) or mean squared difference (Cohen, 1968), is larger when the variables differ more. Given the relationship of quadratically weighted kappa with correlation and intraclass correlation (Fleiss & Cohen, 1973), it is reasonable to employ this member of the kappa family for numerical variables, unless a good reason exists to use linearly weighted kappa. No other member of the kappa family appears to be commonly used. For reasons that perhaps reflect the original discussion of weighted kappa (Cohen, 1968), the literature related to kappa has been confined to variables that assume only a finite number of values. The discussion in this report makes no such restriction.

Measures of prediction accuracy measure the ability to predict one variable by use of one or more other variables. The predicted variable and the predicting variables can be numerical measures or nominal variables. No symmetry is involved. Prediction of a human holistic score by use of computer-generated variables can be considered even though one score is predicted by many computer-generated variables and the human score has very different measurement properties than do computer-generated variables. The principle of proportional reduction of error applies again, and mean squared error remains quite important, given its connection to least squares. In the case of numerical predicted variables and the use of mean squared error, proportional reduction of error leads to the coefficient of determination (Wright, 1920; Yule, 1897). Mean absolute error has been considered for measurement of prediction (Haberman, 1996, pp. 397–400); however, its use has not been common despite a literature concerning the use of least absolute deviations instead of least squares for linear models (Bloomfield & Steiger, 1983). In the case of nominal dependent variables, probability prediction can be used to provide flexible selection of measures (Gilula & Haberman, 1995a, 1995b; Haberman, 1982a, 1982b). Of the common options, the tau measure (Goodman & Kruskal, 1954) has the advantage that it is not 0 if the predicted variable is not independent of the predictor or predictors. The original lambda measure for prediction (Goodman & Kruskal, 1954) can be 0 even if the predicted variable is rather strongly related to the predictor or predictor. This problem arises because this lambda compares conditional and unconditional classification errors, a criterion that is highly insensitive in many practical situations. As a consequence, it is recommended that the lambda measure for prediction be deprecated, despite its frequent use.

Owing to the difference in the purposes of measures of agreement and measures of prediction, predictors should not be evaluated in terms of measures of agreement, and measures of agreement should not be evaluated in terms of measures of prediction.

Measures of agreement are readily estimated from random samples under very general conditions. Measures of prediction accuracy may be easily estimated from such samples in simple cases, but they are very difficult to estimate in general cases without model assumptions. It is very important to understand that the assumption of simple random sampling is not necessarily valid in applications. This issue is especially important in educational testing in survey assessments, cases in which multiple administrations are treated together, and cases in which multiple prompts are treated together. As is the case for any estimation problem, failure to account for the sampling procedure employed can result in substantial errors in evaluation of the accuracy of estimates.

Measures of Agreement

As previously noted, a measure of agreement assesses the extent to which two or more variables can be regarded as interchangeable, so that the different measures can be substituted for each other. To begin, let $J \geq 2$ be an integer, and let $\mathbf{X}$ be the J-dimensional random vector on a population S with elements X_j, $1 \leq j \leq J$. If $J = 2$, then $\mathbf{X}$ may sometimes be written as (X_1, X_2), while for $J = 3$, the notation (X_1, X_2, X_3) may sometimes be used. Measures of agreement have basic symmetry properties. For example, a measure for (X_1, X_2) should be the same as a measure for (X_2, X_1). In general, let $\boldsymbol{\Omega}_J$ be the set of permutations on the set $\bar{J}$ of positive integers no greater than J, so that ω is in $\boldsymbol{\Omega}_J$ if and only if ω is a function from $\bar{J}$ onto $\bar{J}$. For ω in $\boldsymbol{\Omega}_J$, let $\mathbf{X}_\omega$ denote the J-dimensional random vector with element j equal to $X_{\omega(j)}$ for $1 \le j \le J$. For $J = 2$, $\omega(1) = 2$, and $\omega(2) = 1$, $\mathbf{X}_\omega = (X_2, X_1)$. A measure of agreement for the random vector $\mathbf{X}$ has the symmetry property that the measure for $\mathbf{X}$ has the same value as the measure for $\mathbf{X}_\omega$ for all permutations ω in $\boldsymbol{\Omega}_J$. For example, the X_j, $1 \le j \le J$, might be the numerical ratings assigned by J rating systems to a population S of essays, so that, for an essay s in the population S, $X_j(s)$, $1 \le j \le J$, is the numerical essay rating assigned to s under system j. The variables X_j, $1 \le j \le J$, could also correspond to J thermometers or J procedures for detection of a disease. The case of $J = 2$ variables is the most familiar one. In both discussions of estimation and development of the kappa measures, mutually independent random vectors $\mathbf{X}_i$, $i \ge 1$, are considered with the same distribution as $\mathbf{X}$. The elements of $\mathbf{X}_i$ are X_{ij}, $1 \le j \le J$. An important special case of agreement measurement involves symmetric random vectors. Here $\mathbf{X}$ is symmetric if $\mathbf{X}_\omega$ has the same distribution for all permutations ω in $\boldsymbol{\Omega}$. This case is important in psychometrics in the study of parallel-form reliability and rater reliability (Lord & Novick, 1968); however, agreement often must be evaluated when distributions are not symmetric. For example, agreement of two thermometers is still of interest if the thermometers are not perfectly calibrated.

Discrepancy Functions

A measure of agreement may assess absolute or relative agreement. A general description of measures of agreement may be based on the notion of discrepancy functions. Here a discrepancy function d is a real convex nonnegative symmetric on the set R of real numbers such that $d(z) = 0$ if and only if the real number z is 0. A convexity assumption holds if

$$\frac{d(y) - d(x)}{y - x} \le \frac{d(z) - d(x)}{z - x}$$

for all real numbers x, y, and z such that $x < y < z$. The symmetry condition holds if $d(z) = d(-z)$ for all real z. The discrepancy of the numbers x and y is then $d(y-x)$, so that $d(y-x) = 0$ if and only if $y = x$, as is the case if x and y are not discrepant, and the discrepancy $d(y-x)$ of x and y is the same as the discrepancy $d(x-y)$ of y and x. The convexity condition and the requirement that $d(0) = 0$ imply that $d(z)/z$ is nondecreasing for $z > 0$ and $d(z)$ approaches ∞ if $|z|$ approaches ∞. Thus, for any real numbers w, x, y, and z such that $|x - w| < |z - y|$, $d(x-w) \le (|x-w|/|z-y|)d(z-y)$, so that the discrepancy of x and w is no greater than the discrepancy of y and z. For convenience, it is assumed that $d(1) = 1$, so that $d(z) \le |z|$ if $|z| \ge 1$ and $d(z) \ge |z|$ if $|z| \le 1$. For further convenience, it is assumed that, for a real number $h(d) \ge 1$, $d(z)/|z|^{h(d)}$ converges to a positive real number if z approaches ∞. This assumption implies that $d(Z)$ has a finite expectation for a real random variable Z if and only if $|Z|^{h(d)}$ has a finite expectation. It is helpful but not necessary that d be strictly convex, so that

$$\frac{d(y) - d(x)}{y - x} < \frac{d(z) - d(x)}{z - x}$$

for real numbers x, y, and z such that $x < y < z$. For a detailed discussion of the properties of convex functions used in this report, see Rockafellar (1970). For a general treatment of convex functions in estimation, see Haberman (1989). The definition of discrepancy function used here substantially simplifies conditions for measures to be defined and for results to hold, and it does permit inclusion of the most common measures for numerical variables. As discussed in the section Nominal Variables, the discrepancy function also applies to the study of nominal variables.

The quadratic discrepancy function q is the infinitely differentiable strictly convex function that satisfies $q(z) = z^2$ for real z. Here $h(q) = 2$. The function q is the most appropriate discrepancy function for most purposes. It is closely related to standard statistical concepts, such as least squares, variance, correlation, and mean squared error. The other most important discrepancy function is the linear discrepancy function a such that $a(z) = |z|$ for real z. Here $h(a) = 1$. The linear discrepancy function is associated with statistical measures, such as mean absolute deviations and least absolute deviations (Bloomfield & Steiger, 1983; Cicchetti & Allison, 1971). The function a is not differentiable at 0 and is not strictly convex, for $[a(y) - a(x)]/(y - x)$ is the same as $[a(z) - a(x)]/(z - x)$ if $0 \le x < y < z$.

The absolute agreement measure uses the nonnegative convex function d_J on the set R^J of J-dimensional vectors to measure average pairwise discrepancy (Conger, 1980; Hoeffding, 1948). For $\mathbf{x}$ in R^J with elements x_j, $1 \le j \le J$,

$$d_J(\mathbf{x}) = \frac{2}{J(J-1)} \sum_{j=2}^{J} \sum_{k=1}^{j-1} d\left(x_j - x_k\right)$$

is the average value of the discrepancy $d(x_j - x_k)$ of x_j and x_k for $1 \le k < j \le J$. If $J = 2$, then for $\mathbf{x} = (x_1, x_2)$,

$$d_2\left(\left(x_1, x_2\right)\right) = d_2\left(\left(x_2, x_1\right)\right) = d\left(x_2 - x_1\right).$$

If $J = 3$ and $\mathbf{x} = (x_1, x_2, x_3)$,

$$d_J\left(\left(x_1, x_2, x_3\right)\right) = \frac{1}{3}\left[d\left(x_2 - x_1\right) + d\left(x_3 - x_1\right) + d\left(x_3 - x_2\right)\right].$$

The function $d_J(\mathbf{x}) = 0$ if and only if $x_j = x_1$ for $2 \le j \le J$. The function d_J is symmetric, for $d_J(\mathbf{x}) = d_J(\mathbf{x}_\omega)$ for all $\mathbf{x}$ in R^J and ω in Ω_J. For example, $d_3((x_1, x_2, x_3)) = d_3((x_2, x_1, x_3)) = d_3((x_3, x_1, x_2)) = d_3((x_3, x_2, x_1)) = d_3((x_1, x_3, x_2)) = d_3((x_2, x_3, x_1))$.

To simplify discussion, let $L_{h(d)}$ be the linear space of real random variables Z such that $|Z|^{h(d)}$ has a finite expectation, and assume that X_j is in $L_{h(d)}$ for $1 \le j \le J$. The absolute agreement measure $D_J(\mathbf{X}; d) = E(d_J(\mathbf{X})) \ge 0$ (Conger, 1980) is a symmetric finite function of $\mathbf{X}$. The expected discrepancy $D_J(\mathbf{X}; d) = 0$ if and only if $X_j = X_1$ with probability 1 for $2 \le j \le J$.

In the case of quadratic discrepancy, because $h(q) = 2$, it is assumed that $E\left(X_j^2\right)$ is finite for $1 \le j \le J$, so that the expectation $E(X_j)$ of X_j, the variance $\sigma^2(X_j) = E([X_j - E(X_j)]^2)$ of X_j, and the standard deviation $\sigma(X_j)$ of X_j are finite. If $\mathbf{x}$ is a J-dimensional vector with elements x_j, $1 \le j \le J$, let $m_J(\mathbf{x})$ be the sample mean $J^{-1} \sum_{j=1}^{J} x_j$ of the elements of $\mathbf{x}$, and let $s_J^2(\mathbf{x})$ be the sample variance be the sample variance $(J-1)^{-1} \sum_{j=1}^{J} [x_j - m_J(\mathbf{x})]^2$ of the elements of $\mathbf{x}$. Then $d_J(\mathbf{x})$ is

$$q_J(\mathbf{x}) = \frac{2}{J(J-1)} \sum_{j=2}^{J} \sum_{k=1}^{j-1} \left(X_j - X_k\right)^2 = 2s_J^2(\mathbf{x}),$$

and

$$D_J\left(\mathbf{X}; q\right) = \frac{2}{J(J-1)} \sum_{j=2}^{J} \sum_{k=1}^{j-1} E\left(\left[X_j - X_k\right]^2\right) = E\left(q_J(\mathbf{X})\right) = 2E\left(s_J^2(\mathbf{X})\right)$$

(Hoeffding, 1948).

In the case of absolute discrepancy, because $h(a) = 1$, it is assumed that $E(|X_j|)$ is finite for $1 \le j \le J$. Here $d_J(\mathbf{x}) = a_J(\mathbf{x})$, where $a_J(\mathbf{x})$ is the average absolute difference $|x_j - x_k|$ for $1 \le k < j \le J$, and $D_J(\mathbf{X}; a) = E(a_J(\mathbf{X}))$ is the average mean absolute difference $E(|X_j - X_k|)$ for $1 \le k < j \le J$. The Cauchy–Schwarz inequality implies that $[D_J(\mathbf{X}; a)]^2 \le E([a_J(\mathbf{X})]^2) \le D_J(\mathbf{X}; q)$ when $E\left(X_j^2\right)$ is finite for $1 \le j \le J$. For $J = 2$ variables, $D_2([X_1, X_2]; q) = E([X_2 - X_1]^2)$ and $D_2([X_1, X_2]; a) = E(|X_2 - X_1|)$. In some cases, use of sorting facilitates computation. Let $\Omega_J(\mathbf{x})$ be the set of permutations ω in Ω such that $x_{\omega(j)}$ is nondecreasing for $1 \le j \le J$. For any positive integer $j \le J$, $x_{\omega(j)}$ has a constant value $o_{Jj}(\mathbf{x})$ for ω in $\Omega_J(\mathbf{x})$, and $a_J(\mathbf{x})$ is the average of $2(J-1)^{-1}(2j - J - 1)o_{Jj}(\mathbf{x})$ for $1 \le j \le J$ (Haberman, 1996, p. 395). If $J = 2$,

$$a_2(\mathbf{x}) = o_{22}(\mathbf{x}) - o_{21}(\mathbf{x})$$

is the difference between the largest and smallest elements of $\mathbf{x}$. For $J = 3$,

$$a_3(\mathbf{x}) = 2\left[o_{33}(\mathbf{x}) - o_{31}(\mathbf{x})\right]/3,$$

while for $J = 4$,

$$a_4(\mathbf{x}) = \left[3o_{44}(\mathbf{x}) + o_{43}(\mathbf{x}) - o_{42}(\mathbf{x}) - 3o_{14}(\mathbf{x})\right]/6.$$

For a positive integer $j \le J$, the less precise notation $x_{(j)}$ is often employed for $o_{Jj}(\mathbf{x})$, and $X_{(j)} = o_{Jj}(\mathbf{X})$ is then the jth-order statistic of the elements of $\mathbf{X}$.

For quadratic discrepancy, let $\mathrm{Cov}(X_j, X_k)$ denote the covariance of X_j and X_k for positive integers j and k no greater than J, and let $\rho(X_j, X_k)$ denote the correlation of X_j and X_k when X_j and X_k have positive variance. Then

$$D_J\left(\mathbf{X}; q\right) = 2s_J^2\left(E\left(\mathbf{X}\right)\right) + 2E\left(s_J^2\left(\mathbf{X} - E\left(\mathbf{X}\right)\right)\right).$$

If $J = 2$,

$$D_2\left(\left(X_1, X_2\right); q\right) = \left[E\left(X_2\right) - E\left(X_1\right)\right]^2 + \sigma^2\left(X_1\right) + \sigma^2\left(X_2\right) - 2\mathrm{Cov}\left(X_1, X_2\right),$$

where $\mathrm{Cov}(X_1, X_2) = \sigma(X_1)\sigma(X_2)\rho(X_1, X_2)$ if X_1 and X_2 have positive variance.

If $\mathbf{X}$ is symmetric, $D_J(\mathbf{X}; d) = D_2((X_1, X_2); d) = E(d(X_2 - X_1))$. In the case of quadratic discrepancy, if the condition is added that X_1 has positive variance, then $D_J(\mathbf{X}; q) = 2\sigma^2(X_1)[1 - \rho(X_1, X_2)]$. Under the further condition that X_1 and X_2 have a joint bivariate normal distribution, $D_J(\mathbf{X}; a) = (2/\pi^{1/2})\sigma(X_1)[1 - \rho(X_1, X_2)]^{1/2}$.

Measures of absolute agreement do not consider whether the variables X_j, $1 \le j \le J$, exhibit a small expected discrepancy because they are highly related or whether the small expected discrepancy arises because the variables do not vary. For example, consider quadratic discrepancy. Let $J = 2$. Consider the symmetric case with X_1 and X_2 with the same distribution and X_1 has positive variance. Then $D_2(\mathbf{X}; q)$ can be small if $\sigma^2(X_1)$ is small, even if X_1 and X_2 are uncorrelated. On the other hand, $D_2(\mathbf{X}; q)$ can also be small if the variance $\sigma^2(X_1)$ is large but the correlation $\rho(X_1, X_2)$ is close to 1.

Nominal Variables

In the case of nominal variables, a different approach is appropriate. Here a restricted version of the measures in Gilula and Haberman (1995a) provides the basis for analysis. Let X_j, $1 \le j \le J$, be in $\overline{r}$. For real y, let δ_y be the real function on the real line with value $\delta_y(x)$ at real x, where $\delta_y(x) = 1$ if $y = x$ and $\delta_y(x) = 0$ if $y \ne x$. If $\mathbf{x}$ in R^J has elements x_j, $1 \le j \le J$, let $\boldsymbol{\delta}_{Jy}(\mathbf{x})$ be the J-dimensional vector with elements $\delta_y(x_j)$, $1 \le j \le J$. Let $d_{Jr}\left(\mathbf{x}\right) = \sum_{j=1}^{r} d_J\left(\boldsymbol{\delta}_r\left(\mathbf{x}\right)\right)$ for any J-dimensional vector $\mathbf{x}$, and let $D_{Jr}(\mathbf{X}) = E(d_{Jr}(\mathbf{X}))$. For any discrepancy function d, $d_{Jr}(\mathbf{x})$ is twice the fraction of positive integers j and k such that $k < j \le J$ and $x_j \ne x_k$, and $D_{Jr}(\mathbf{X})$ is the average of $2P(X_j \ne X_k)$ for positive integers j and k such that $k < j \le J$, where $P(X_j \ne X_k)$ is the probability that X_j and X_k are unequal. Thus $D_{Jr}(\mathbf{X})$ is nonnegative and no greater than 2, and $D_{Jr}(\mathbf{X}) = 0$ if and only if x in $\overline{r}$ exists such that $X_j = x$ with probability 1 for $1 \le j \le J$. The measure $D_{Jr}(\mathbf{X})$ is twice the expected fraction of pairs X_j and X_k that are discrepant. In addition, $D_{Jr}(\mathbf{X}) = D_{Js}(\mathbf{X})$ if $s \ge r$ is an integer. If $r = 2$, then $D_{J2}(\mathbf{X}) = 2D_J(\mathbf{X}; d)$. The function $D_{Jr}(\mathbf{X}; d)$ is a symmetric function of $\mathbf{X}$. The function $D_{Jr}(\mathbf{X})$ is also symmetric over permutation transformations in the sense that $D_{Jr}(\mathbf{X}) = D_{Jr}(\upsilon(\mathbf{X}))$, where $\upsilon(\mathbf{X})$ has elements $\upsilon(X_j)$, $1 \le j \le J$, and υ is in the set Ω_r of permutations on $\overline{r}$. Thus, if languages are Japanese, Chinese, and Korean, and the X_j are language classifications with integer values 1, 2, and 3, then which numerical code from 1 to 3 is assigned to which language has no impact on the measure. If $J = 2$ or $\mathbf{X}$ is symmetric, then $D_{Jr}(\mathbf{X}) = 2P(X_2 \ne X_1)$.

An alternative formula for $D_{Jr}(\mathbf{X})$ is worth noting. For any positive integer $y \le r$, $m_J(\boldsymbol{\delta}_{Jy}(\mathbf{x}))$ is the fraction of positive integers $j \le J$ such that $x_j = y$, and the sum $\sum_{y=1}^{r} m_J\left(\boldsymbol{\delta}_{Jy}\left(\mathbf{x}\right)\right) = 1$ if x_j is in $\overline{r}$ for all positive integers $j \le J$. The fraction of distinct positive integers j and k no greater than J such that $x_j \ne x_k$ is

$$D_{Jr}\left(\mathbf{X}\right) = \frac{2J}{J-1}\left\{1 - \sum_{y=1}^{r} E\left(\left[m_J\left(\boldsymbol{\delta}_{Jy}\left(\mathbf{X}\right)\right)\right]^2\right)\right\},$$

so that

$$D_{Jr}\left(\mathbf{X}\right) = \frac{2J}{J-1}\left\{1 - \sum_{y=1}^{r} E\left(\left[m_J\left(\boldsymbol{\delta}_{Jy}\left(\mathbf{X}\right)\right)\right]^2\right)\right\}.$$

In studying nominal random variables, it is important to understand the strong constraints on the probability $P(\mathbf{X} = \mathbf{x})$ that $\mathbf{X} = \mathbf{x}$, where $\mathbf{x}$ is a J-dimensional vector with all elements in $\overline{r}$, by the marginal probabilities $P(X_j = x)$ that $X_j = x$, $1 \le j \le J$, $1 \le x \le r$. Consider Table 1 for the case of $J = 2$ and $r = 2$. In this case, the probability $P(X_2 \ne X_1) = .2$ that X_1 and X_2 are not equal is relatively low; however, $P(X_2 \ne X_1) \le .2$ whenever $P(X_1 = 1) = P(X_2 = 1) = .9$ and $P(X_1 = 2) = P(X_2 = 2) = .1$, so that the low value of $P(X_2 \ne X_1)$ just reflects the marginal distributions of X_1 and X_2.

Table 1 An Example of $P(\mathbf{X}=\mathbf{x})$ for $J=2$ and $r=2$

x_1	$x_2=1$	$x_2=2$
1	.8	.1
2	.1	0

Note. The elements of $\mathbf{x}$ are x_1 and x_2.

Given limitations associated with absolute measures of discrepancy, it is appropriate to consider proportional reduction of expected discrepancy (Costner, 1965). With this approach, an absolute agreement measure such as $D(\mathbf{X}; q)$ is compared to one or more relevant alternative agreement measures. The two cases considered here are kappa measures (Cohen, 1960, 1968) and lambda measures for agreement (Goodman & Kruskal, 1954). Kappa measures are more commonly used, but lambda measures for agreement are also quite appropriate to consider. The case of $J=2$ is the most commonly encountered. The reliability terminology in Goodman and Kruskal (1954) is avoided to avoid conflict with conventional psychometric concepts of reliability (Lord & Novick, 1968).

Kappa Measures

In the case of kappa measures of agreement for numerical measures, $D_J(\mathbf{X}; d)$ is compared to $D_J(\mathbf{X_J}; d)$, where $\mathbf{X_J}$ is the J-dimensional random vector with elements X_{jj}, $1 \le j \le J$. This random variable is significant because the elements X_{jj} of $\mathbf{X_J}$, $1 \le j \le J$, are mutually independent and X_{jj} and X_j have the same distribution. Thus the agreement measured by $D_J(\mathbf{X}; d)$ is compared to the agreement measured by a random vector with elements that are mutually independent but have the same respective marginal distributions as the corresponding elements of $\mathbf{X}$. Despite its appearance, $D_J(\mathbf{X_J}; d)$ can be evaluated by use of only the independent random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$, which both have the same distribution as $\mathbf{X}$, for $D_J(\mathbf{X_J}; d)$ is the expectation of $e_J(\mathbf{X}_1, \mathbf{X}_2; d)$, where, for the K-dimensional vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ with respective elements x_{j1}, $1 \le j \le J$, and x_{j2}, $1 \le j \le J$, $e_J(\mathbf{x}_1, \mathbf{x}_2)$ is the average of $d(x_{j1} - x_{k2})$ for all distinct positive integers j and k such that $k < j \le J$. For reasons that are not obvious, κ measures have been used traditionally only for cases in which the X_j, $1 \le j \le J$, are restricted to a finite set of possible values (Cohen, 1960, 1968; Conger, 1980). Despite this tradition, no reason exists not to provide all definitions and results for general random vectors $\mathbf{X}$.

An alternative approach to $D_J(\mathbf{X_J}, d)$ may be based on iterated expectations. With this approach, explicit reference to $\mathbf{X}_J$ does not appear in formulas. To apply this approach, for a real random variable Z in $L_{h(d)}$, let the nonnegative finite convex function $G(Z; d)$ be the function on the real line with value $G(z, Z; d) = E(d(Z - z))$ at real z. This function will be used for a variety of applications in this report. Because d is convex, $G(Z; d)$ is also convex. Because $d(Z - z)$ approaches ∞ as $|z|$ approaches ∞, $G(z, Z; d)$ approaches ∞ as $|z|$ approaches ∞. Therefore a closed and bounded interval $C(Z; d) = [c_L(Z; d), c_U(Z; d)]$ with midpoint $c(Z; d) = [c_L(Z; d) + c_U(Z; d)]/2$ exists such that $G(z, Z; d)$ equals the infimum of $G(Z; d)$ if and only if z is in $C(Z; d)$ (Rockafellar, 1970, p. 264). If d is strictly convex, then $c(Z; d) = c_L(Z; d) = c_U(Z; d)$ is the only member of $C(Z; d)$.

To apply iterated expectations, for real random variables Z_1 and Z_2 in $L_{h(d)}$, let $G(Z_1, Z_2; d)$ be the finite random variable with value $G(z, Z_2; d)$ if $Z_1 = z$ for the real number z. If Z_1 and Z_2 are independent, then Fubini's theorem implies that $E(G(Z_1, Z_2; d)) = E(G(Z_2, Z_1; d)) = E(d(Z_1 - Z_2))$ is finite. Thus $D_J(\mathbf{X_J}; d)$ is the average of $E(G(X_k, X_j; d))$ for $1 \le k < j \le J$. The mean discrepancy $D_J(\mathbf{X_J}; d) = 0$ if and only if some real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$. If no real x exists such that $X_j = x$ with probability 1, then

$$\kappa_J(\mathbf{X}; d) = \frac{D_J\left(\mathbf{X_J}; d\right) - D_J(\mathbf{X}; d)}{D_J\left(\mathbf{X_J}; d\right)},$$

so that $\kappa_J(\mathbf{X}; d) = \kappa(\mathbf{X}_\omega; d) \le 1$ for each permutation ω in Ω_J, with $\kappa_J(\mathbf{X}; d) = 1$ if and only if $X_j = X_1$ with probability 1 for $2 \le j \le J$. In addition, $\kappa_J(\mathbf{X}; d) = 0$ if the X_j, $1 \le j \le J$, are pairwise independent (Cohen, 1968), although $\kappa_J(\mathbf{X}; d) = 0$ does not imply that the X_j, $1 \le j \le J$, are pairwise independent. If $\mathbf{X}$ is symmetric and no real x exists such that $X_1 = x$ with probability 1, then

$$\kappa_J(\mathbf{X}; d) = 1 - \frac{E\left(d\left(X_2 - X_1\right)\right)}{E\left(G\left(X_1, X_2; d\right)\right)}.$$

Quadratically weighted kappa is an especially attractive case due to its relationship to correlation and variance. If no real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$, then

$$D_J\left(\mathbf{X_J}; q\right) = 2s_J^2\left(E\left(\mathbf{X}\right)\right) + 2m_J\left(\sigma^2\left(\mathbf{X}\right)\right),$$

where $\sigma^2(\mathbf{X})$ is the J-dimensional vector with elements $\sigma^2(X_j)$ for $1 \le j \le J$, and

$$\kappa_J\left(\mathbf{X}; q\right) = \frac{m_J\left(\sigma^2\left(\mathbf{X}\right)\right) - E\left(s_J^2\left(\mathbf{X} - E\left(\mathbf{X}\right)\right)\right)}{s_J^2\left(E\left(\mathbf{X}\right)\right) + m_J\left(\sigma^2\left(\mathbf{X}\right)\right)}.$$

The numerator is also the average $(J-1)^{-1}[J\sigma^2(m_J(\mathbf{X})) - m_J(\sigma^2(\mathbf{X}))]$ of $\mathrm{Cov}(X_j, X_k)$ for $1 \le k < j \le J$, so that $\kappa_J(\mathbf{X}; q) \ge -(J-1)^{-1}$. If $J = 2$, X_1 and X_2 have finite variances, and no real x exists such that $X_1 = x$ and $X_2 = x$ with probability 1, then

$$\kappa_2\left(\mathbf{X}; q\right) = \frac{2\mathrm{Cov}\left(X_1, X_2\right)}{\left[E\left(X_2\right) - E\left(X_1\right)\right]^2 + \sigma^2\left(X_1\right) + \sigma^2\left(X_2\right)}.$$

In this case, $\kappa_2(\mathbf{X}; q) = -1$ if $E(X_1) = E(X_2)$ and $X_1 + X_2 = 2E(X_1)$ with probability 1. In the symmetric case, consider $\sigma^2(X_1) > 0$. Then $\kappa_J(\mathbf{X}; q) = \rho(X_1, X_2)$ is the intraclass correlation for exchangeable raters (Fleiss & Cohen, 1973). It is also the parallel-form reliability in psychometrics if X_1 and X_2 are results of parallel tests. In general, if $J = 2$ and $\sigma^2(X_1)$ and $\sigma^2(X_2)$ are positive, then $\kappa_2(\mathbf{X}; q)$ and the correlation $\rho(X_1, X_2)$ have the same sign and $|\kappa_2(\mathbf{X}; q)| \le |\rho(X_1, X_2)| \le 1$.

Linear transformations of the X_j, $1 \le j \le J$, can affect quadratically weighted kappa. For example, if $J = 2$, X_1, and X_2 have positive variances, and $Z = E(X_2) + [\sigma(X_2)/\sigma(X_1)][X_1 - E(X_1)]$, then $\kappa_2((Z, X_2); q) = \rho(X_1, X_2)$. Unless $E(X_1) = E(X_2)$ and $\sigma^2(X_1) = \sigma^2(X_2)$, $\kappa_2((Z, X_2); q) > \kappa_2(\mathbf{X}; q)$; however, linear transformations of this kind are only appropriate in limited circumstances. For example, if X_1 measures temperature in Celsius and X_2 provides Fahrenheit measures, then use of $Z = 32 + 1.8X_1$ rather than X_1 is appropriate for a sensible comparison of the thermometers. Nonetheless, consistent linear transformations on all the X_j do not affect quadratically weighted kappa. If $Z_j = c + bX_j$ for $1 \le j \le J$, c is a real constant, b is a positive real constant, and $\mathbf{Z}$ is the J-dimensional vector with elements Z_j for $1 \le j \le J$, then $\kappa(\mathbf{Z}; q) = \kappa(\mathbf{X}; q)$. This result does not hold for general monotone transformations. Thus $\kappa(g(\mathbf{X}); q)$ may differ from $\kappa(\mathbf{X}; q)$ if g is a strictly increasing real function on the real line, $g(\mathbf{X})$ is the J-dimensional random vector with elements $g(X_j)$ for $1 \le j \le J$, and each $g(X_j)$ is in L_2. This report takes the values of the X_j as given and does not consider nonlinear transformations.

Linearly weighted kappa corresponds to the absolute discrepancy measure a, so that $\kappa_J(\mathbf{X}; a) = 1 - E(a_J(\mathbf{X}))/E(a_J(\mathbf{X_J}))$ if no real number x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$. As in the case of quadratically weighted kappa, $\kappa(\mathbf{Z}; a) = \kappa(\mathbf{X}; a)$. If $\mathbf{X}$ is symmetric and has a multivariate normal distribution with $\sigma^2(X_1) > 0$, then $\kappa(\mathbf{X}; a) = 1 - [1 - \rho(X_1, X_2)]^{1/2}$. Thus $\kappa(\mathbf{X}; a)\{1 + [1 - \rho(X_1, X_2)]^{1/2}\} = \rho(X_1, X_2)$, $\kappa(\mathbf{X}; a)$, and $\rho(X_1, X_2)$ have the same sign, and $|\kappa(X; a)| \le |\kappa(X; q)| = |\rho(X_1, X_2)|$.

Kappa Measures and Nominal Variables

In the case of nominal random variables, let r be a positive integer, and let X_j be in $\overline{r}$ for $1 \le j \le J$. Assume that no y in $\overline{r}$ exists such that $X_j = y$ with probability 1 for $1 \le j \le J$. Then $\kappa_{Jr}(\mathbf{X}) = [D_{Jr}(\mathbf{X_J}) - D_{Jr}(\mathbf{X})]/D_{Jr}(\mathbf{X_J})$, so that $\kappa_{Jr}(\mathbf{X})$ is a symmetric function of $\mathbf{X}$ and $\kappa_{Jr}(\mathbf{X})$ is also symmetric with respect to permutations of the values of the X_j. In the case of $J = 2$, $\kappa_{2r}(\mathbf{X})$ is the original kappa measure (Cohen, 1960). In computations, $D_{Jr}(\mathbf{X_J})$ is twice the average of the probabilities

$$P\left(X_{jj} \ne X_{kk}\right) = 1 - \sum_{y=1}^{r} P\left(X_j = y\right) P\left(X_k = y\right)$$

for $1 \le k < j \le J$. Let $\mathbf{p}_J(\mathbf{X})$ be the r-dimensional vector with elements $p_{Jy}(\mathbf{X})$, $1 \le y \le r$, equal to the average of the probabilities $P(X_j = y)$ for $1 \le j \le J$. Thus $\mathbf{p}_J(\mathbf{X_J}) = \mathbf{p}_J(\mathbf{X})$. It follows that

$$\frac{1}{2}D_{Jr}\left(\mathbf{X_J}\right) = 1 - \frac{J}{J-1}\sum_{y=1}^{r}\left[p_{Jy}\left(\mathbf{X}\right)\right]^2 + [J(J-1)]^{-2}\sum_{y=1}^{r}\sum_{j=1}^{J}\left[P\left(X_j = y\right)\right]^2.$$

The equation $\kappa_{Jr}(\mathbf{X}) = 1$ can only hold if $X_j = X_1$ with probability 1. If the X_j, $1 \le j \le k$, are pairwise independent, then $\kappa_{J2}(\mathbf{X}) = 0$. If $r = 2$, then $\kappa_{J2}(\mathbf{X}) = \kappa_J(\mathbf{X}; d)$. For example, in Table 1, $\kappa_{22}(\mathbf{X}) = \kappa_2(\mathbf{X}; q) = -1/9$ indicates that the agreement of X_1 and X_2 is less than would be expected were X_1 and X_2 independent.

Lambda Measures for Agreement

Lambda measures for agreement provide a case of proportional reduction of error that does not involve use of the hypothetical vector $\mathbf{X}_J$. They generalize a measure proposed in Goodman and Kruskal (1954). Because they have received relatively little attention despite the importance of their source, details here are generally new, although many results are related to those in Haberman (1989). As evident from results to be presented, their use may also have been affected by the more conservative picture of agreement they provide relative to the picture provided by kappa.

For lambda measures for agreement, the base for comparison involves prediction by a constant. For real x and a J-dimensional vector $\mathbf{x}$ with elements x_j for $1 \le j \le J$, let $\eta_J(\mathbf{x}; d)$ be the nonnegative real convex function with value $\eta_J(x, \mathbf{x}; d) = J^{-1} \sum_{j=1}^{J} d\left(x_j - x\right)$ at x. Let X_j be in $L_{h(d)}$ for $1 \le j \le J$, and let $G_J(\mathbf{X}; d)$ be the finite and nonnegative convex function $E(\eta_J(\mathbf{X}; d))$. Let $G_J(\mathbf{X}; d)$ have value $G(x, \mathbf{X}; d)$ at x. Let $G_{J-}(\mathbf{X}; d)$ be the infimum of $G_J(\mathbf{X}; d)$, so that $G_J(\mathbf{X}; d)$ is positive if and only if no real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$. As in the case of $G(Z; d)$ for Z in $L_{h(d)}$, $G_J(x, \mathbf{X}; d)$ approaches ∞ as $|x|$ approaches ∞, a nonempty closed and bounded interval $C_J(\mathbf{X}; d) = [c_{JL}(\mathbf{X}; d), c_{JU}(\mathbf{X}; d)]$ exists with midpoint $c_J(\mathbf{X}; d) = [c_{JL}(\mathbf{X}; d) + c_{JU}(\mathbf{X}; d)]2$ such that real x satisfies $G_J(x, \mathbf{X}; d) = G_{J-}(\mathbf{X}; d)$ if and only if x is in $C_J(\mathbf{X}; d)$. If d is strictly convex, then $c_J(\mathbf{X}; d) = c_{JL}(\mathbf{X}; d) = c_{JU}(\mathbf{X}; d)$ is the only element of $C_J(\mathbf{X}; d)$. If no real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$, then $G_{J-}(\mathbf{X}; d) > 0$, and the lamba measure for agreement

$$\lambda_{Ja}(\mathbf{X}; d) = \frac{G_{J-}(\mathbf{X}; d) - D_J(\mathbf{X}; d)}{G_{J-}(\mathbf{X}; d)}$$

is no greater than 1 and is equal to 1 if and only if $X_j = X_1$ with probability 1 for $2 \le j \le J$. The symmetry property holds that $\lambda_{Ja}(\mathbf{X}_\omega; d) = \lambda_{Ja}(\mathbf{X}; d)$ for each permutation ω in Ω_J.

To interpret the lambda measure, observe that the equation $D_J(\mathbf{X}) = E(d_J(\mathbf{X}))$ and the symmetry of d imply that

$$d_J(\mathbf{X}) = \frac{1}{J(J-1)} \sum_{j=2}^{J} \sum_{k=1}^{j-1} \left[d\left(X_j - X_k\right) + d\left(X_k - X_j\right)\right].$$

One can regard X_j as a predictor of X_k and X_k as a predictor of X_j if $1 \le k < j \le J$. For a real number x, consider prediction of X_j by x and prediction of X_k by x for $1 \le k < j \le J$. The average discrepancy

$$\frac{1}{J(J-1)} \sum_{j=2}^{J} \sum_{k=1}^{j-1} \left[d\left(X_j - x\right) + d\left(X_k - x\right)\right] = \eta_J(x, \mathbf{X}; d)$$

has expectation equal to $G_J(x, \mathbf{X}; d)$, so that $G_{J-}(\mathbf{X}; d)$ is the minimum achievable expected average from use of x to predict each X_j, $1 \le j \le J$. Thus the comparison of $D_J(\mathbf{X}; d)$ involves a trivial predictor of each X_j that makes no use of any possible differences between the X_j or any possible relationships between the X_j.

Basic properties of $\lambda_{Ja}(\mathbf{X}; d)$ are readily obtained. The measure $\lambda_{Ja}(\mathbf{X}; d) = 1$ if and only if $X_j = X_1$ with probability 1 for $2 \le j \le J$. If each X_j, $1 \le j \le J$, has the same marginal distribution, then $G_J(\mathbf{X}; d) = G(X_1; d)$, $C_J(\mathbf{X}; d) = C(X_1; d)$, $c_{JL}(\mathbf{X}; d) = c_L(X_1; d)$, $c_{JU}(\mathbf{X}; d) = c_U(X_1; d)$, $c_J(\mathbf{X}; d) = c(X_1; d)$, and $G_{J-}(\mathbf{X}; d) = G_-(X_1; d)$.

To compare $\lambda_{Ja}(\mathbf{X}; d)$ and $\kappa_J(\mathbf{X}; d)$, observe that

$$G_{J-}(\mathbf{X}; d) = G_{J-}\left(\mathbf{X}_\mathrm{J}; d\right) \le J^{-1} \sum_{j=1}^{J} E\left(G\left(X_k, X_j; d\right)\right) = J^{-1} \sum_{j=1}^{J} E\left(d\left(X_{jj} - X_{kk}\right)\right)$$

for $1 \le k \le J$ and $1 \le j \le J$ such that $j \ne k$. Averaging over k from 1 to J yields

$$G_{J-}(\mathbf{X}; d) \le \frac{J-1}{J} D_J\left(\mathbf{X}_\mathrm{J}; d\right).$$

Thus $\lambda_{Ja}(\mathbf{X}; d) \le (J-1)^{-1}[J\kappa_J(\mathbf{X}; d) - 1]$. Unless $X_j = X_1$ with probability 1 for $2 \le j \le J$, $\lambda_{Ja}(\mathbf{X}; d) < \kappa_J(\mathbf{X}; d)$. If $\mathbf{X}$ is symmetric, and no real x exists such that $X_1 = x$, then $\lambda_{Ja}(\mathbf{X}; d) = \lambda_{2a}((X_1, X_2); d) \le 2\kappa_2((X_1, X_2); d) - 1 = 2\kappa_J(\mathbf{X}; d) - 1$, so that $\lambda_J(\mathbf{X}; d) < \kappa_J(\mathbf{X}; d)$ unless $X_j = X_1$ with probability 1 for $2 \le j \le J$.

In the case of the quadratic discrepancy function q, each X_j is in L_2,

$$G_J\left(x, \mathbf{X}; q\right) = J^{-1}\sum_{j=1}^{J}\left\{\sigma^2\left(X_j\right) + \left[E\left(X_j\right) - x\right]^2\right\}$$

for real x, and $G_J(y, \mathbf{X}; d) = G_{J-}(\mathbf{X}; q)$ if $x = m_J(E(\mathbf{X}))$ is the average of $E(X_j)$ for $1 \le j \le J$, so that

$$G_{J-}\left(\mathbf{X}; q\right) = m_J\left(\sigma^2\left(\mathbf{X}\right)\right) + \left[(J-1)/J\right] s_J^2\left(E\left(\mathbf{X}\right)\right).$$

If $\mathbf{X}$ is symmetric and X_1 has a positive and finite variance, $\lambda_{Ja}(\mathbf{X}; q) = 2\rho(X_1, X_2) - 1 = 2\kappa_J(\mathbf{X}; q) - 1$. If $\rho(X_1, X_2) < 1/2$, then $\lambda_{Ja}(\mathbf{X}; q)$ is negative, so that, in terms of quadratic discrepancy, X_2 is better predicted by the expectation $E(X_2) = E(X_1)$ than by X_1 as a predictor. For $J = 2$ and X_1 and X_2 parallel tests, $\lambda_{2a}((X_1, X_2)); q) = 2\rho(X_1, X_2) - 1$, where $\rho(X_1, X_2) = \kappa_2((X_1, X_2); q)$ is the parallel-form reliability. This result shows the contrast between the use of reliability by Goodman and Kruskal (1954) for a measure such as $\lambda_{2a}((X_1, X_2); q)$ and the customary psychometric concept of parallel-form reliability.

In the case of the absolute discrepancy function a, $G_{J-}(\mathbf{X}; a)$ is the minimum of $J^{-1}\sum_{j=1}^{J} E\left(|X_j - x|\right)$ for real x. If U is a random variable independent of $\mathbf{X}$ such that $U = j$ with probability J^{-1} for $1 \le j \le J$ and X_U is the random variable with value X_j if $U = j$ and $1 \le j \le J$, then $G_{J-}(\mathbf{X}; a)$ is the mean absolute deviation $\mathrm{MD}(X_U)$ of X_U about the median. Here any member of $C_J(\mathbf{X}; a)$ is a median of X_U. If no real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$, then $\lambda_{Ja}(\mathbf{X}; a)$ is defined. In addition, if each X_j has the same distribution, then $G_{J-}(\mathbf{X}; a)$ is the mean absolute deviation $\mathrm{MD}(X_1)$ of X_1 about the median, where any element of $C(X_1; a)$ is a median of X_1. If $\mathbf{X}$ is symmetric and has a multivariate normal distribution, then

$$\lambda_{Ja}\left(\mathbf{X}; a\right) = 1 - \left\{2\left[1 - \rho\left(X_1, X_2\right)\right]\right\}^{1/2} = 1 - 2^{1/2}\left[1 - \kappa_J\left(\mathbf{X}; a\right)\right],$$

so that $\lambda_{Ja}(\mathbf{X}; a) = \kappa_J(\mathbf{X}; a) = 1$ if $X_1 = X_2$ with probability 1 and $\lambda_{Ja}(\mathbf{X}; a) < \kappa_J(\mathbf{X}; a)$ otherwise. In this case, $\lambda_{Ja}(\mathbf{X}; a)$ is negative if $\rho(X_1, X_2) < 1/2$.

Lambda Measures of Agreement for Nominal Variables

The case of lambda measures of agreement for nominal variables involves probability prediction (Gilula & Haberman, 1995a, 1995b). Let r be a positive integer such that X_j is in $\bar{r}$ for $1 \le j \le J$. Let the unit simplex Π_r be the set of nonnegative r-dimensional vectors with elements with sum 1. For a J-dimensional vector $\mathbf{x}$ with all elements x_j, $1 \le j \le J$, in $\bar{r}$, let $\eta_{Jr}(\mathbf{x}; d)$ be the convex function on Π_r with value $\eta_{Jr}\left(\mathbf{p}, \mathbf{x}; d\right) = \sum_{y=1}^{r}\eta_J\left(p_y, \boldsymbol{\delta}_{Jy}\left(\mathbf{X}\right); d\right)$ at $\mathbf{p}$ in Π_r with elements p_y for y in $\bar{r}$. Let $G_{Jr}(\mathbf{X}; d)$ be the convex function on Π_r with value $G_{Jr}(\mathbf{p}, \mathbf{X}; d) = E(\eta_{Jr}(\mathbf{p}, \mathbf{X}; d))$ at $\mathbf{p}$ in Π_r, and let $G_{Jr-}(\mathbf{X}; d)$ be the infimum of $G_{Jr}(\mathbf{X}; d)$. If no x in $\bar{r}$ exists such that $X_j = x$ with probability 1 for $1 \le j \le r$, then $G_{Jr-}(\mathbf{X}; d)$ is positive and

$$\lambda_{Jar}\left(\mathbf{X}; d\right) = \left[G_{Jr-}\left(\mathbf{Z}; d\right) - D_{Jr}\left(\mathbf{X}\right)\right]/G_{Jr-}\left(\mathbf{X}; d\right).$$

The measure $\lambda_{Jar}(\mathbf{X}; d)$ does not exceed 1 and only equals 1 if $X_j = X_1$ with probability 1 for $2 \le j \le J$. Because Π_r is closed and bounded, a nonempty closed convex subset $C_{Jr}(\mathbf{X}; d)$ of Π_r exists such that $G_{Jr}(\mathbf{p}, \mathbf{X}; d) = G_{Jr-}(\mathbf{X}; d)$ if and only if $\mathbf{p}$ is in $C_{Jr}(\mathbf{X}; d)$. If d is strictly convex, then $C_{Jr}(\mathbf{X}; d)$ has only one member $c_{Jr}(\mathbf{X}; d)$. The use of the notation $\lambda_{Jar}(\mathbf{X}; d)$ reflects the fact that the definition of d affects the lambda measure of agreement. If $r = 2$, then $\lambda_{Ja2}(\mathbf{X}; d) = \lambda_{Ja}(\mathbf{X})$. The measure $\lambda_{Jar}(\mathbf{X})$ is symmetric in $\mathbf{X}$ and in permutations of the values of X_j, $1 \le j \le J$.

In the case of quadratic discrepancy, $G_{Jr}(\mathbf{p}, \mathbf{X}; q)$ is minimized for $\mathbf{p}$ in Π_r if $\mathbf{p} = \mathbf{p}_J(\mathbf{X})$. It follows that

$$\begin{aligned}G_{Jr-}\left(\mathbf{X}; q\right) &= J^{-1}\sum_{y=1}^{r}\sum_{j=1}^{J}\left\{P\left(X_j = y\right)\left[1 - P\left(X_j = y\right)\right] + \left[P\left(X_j = y\right) - p_{Jy}\left(\mathbf{X}\right)\right]^2\right\}\\ &= 1 - \sum_{y=1}^{r}\left[p_{Jy}\left(\mathbf{X}\right)\right]^2.\end{aligned}$$

In the case of absolute discrepancy, $G_{Jr-}(\mathbf{X}; a) = 1 - \max_{y=1}^{r} p_{Jy}(\mathbf{X})$. The original reliability measure (Goodman & Kruskal, 1954) is defined for $J = 2$ as $\lambda_{2ar}(\mathbf{X}; a)$. In Table 1, $\lambda_{2a2}(\mathbf{X}; q) = 2\kappa_{22}(\mathbf{X}) - 1 = -11/9$ and $\lambda_{2a2}(\mathbf{X}; a) = -1$.

Measures of Prediction Accuracy

In measures of prediction accuracy based on the discrepancy function d, the real random variable Y in $L_{h(d)}$ is predicted by a real function $f(\mathbf{Z})$ in $L_{h(d)}$ of a random vector $\mathbf{Z}$ with elements Z_k, $1 \le k \le K < \infty$. To avoid problems with identification, assume that any closed convex set U has a nonempty interior if the probability is 1 that $\mathbf{Z}$ is in U (Berk, 1972). Let $\mathcal{Z}$ be the image of $\mathbf{Z}$, so that a K-dimensional vector $\mathbf{z}$ is in $\mathcal{Z}$ only if $\mathbf{Z}$ can assume the value $\mathbf{z}$. The function f is in a family F of real functions on $\mathcal{Z}$ that includes the family all constant functions on $\mathcal{Z}$. The family F is assumed included in the set $L_{h(d)}(\mathbf{Z})$ of real function f on $\mathcal{Z}$ such that $f(\mathbf{Z})$ is in $L_{h(d)}$. Let $G(Y \mid \mathbf{Z}; d)$ be the convex function on $L_{h(d)}(\mathbf{Z})$ with value $G(f, Y \mid \mathbf{Z}, d) = E(d(Y - f(\mathbf{Z})))$ at f in $L_{h(d)}(\mathbf{Z})$ equal to the expected discrepancy between the predicted value Y and the predictor $f(\mathbf{Z})$. For the family F, $G\_(Y \mid \mathbf{Z}; F, d)$, the infimum of $G(f, Y \mid \mathbf{Z}; d)$ over f in F provides an absolute measure of prediction accuracy. By definition,

$$G\_(Y \mid \mathbf{Z}; d) = G\_\left(Y|\mathbf{Z}; L_{h(d)}(\mathbf{Z}), d\right) \le G\_(Y \mid \mathbf{Z}; F, d) \le G\_(Y; d).$$

If no real number y exists such that $Y = y$ with probability 1, then the corresponding relative measure of prediction accuracy for the family F is

$$\lambda(Y \mid \mathbf{Z}; F, d) = 1 - \frac{G\_(Y \mid \mathbf{Z}; F, d)}{G\_(Y; d)}.$$

It follows that

$$0 \le \lambda(Y \mid \mathbf{Z}; F, d) \le \lambda(Y \mid \mathbf{Z}; d) = \lambda\left(Y|\mathbf{Z}; L_{h(d)}(\mathbf{Z}), d\right) \le 1.$$

If f in F satisfies the condition that $Y = f(\mathbf{Z})$ with probability 1, then $G\_(Y \mid \mathbf{Z}; F, d) = 0$ and $\lambda(Y \mid \mathbf{Z}; F, d) = 1$.

Let $C(Y \mid \mathbf{Z}; F, d)$ be the set of f in F such that $G(f, Y \mid \mathbf{Z}; d) = G\_(Y \mid \mathbf{Z}; d)$. If F is convex and $C(Y \mid \mathbf{Z}; F, d)$ is nonempty, then $C(Y \mid \mathbf{Z}; F, d)$ is convex. Thus $C(Y \mid \mathbf{Z}; d) = C(Y| \mathbf{Z}; L_{h(d)}(\mathbf{Z}))$ is convex if it is nonempty. Use of conditional distributions permits a demonstration that $C(Y \mid \mathbf{Z}; d)$ is not empty. Because Y is a real random variable and $\mathbf{Z}$ is a real random vector, a conditional distribution of Y given $\mathbf{Z}$ can be defined (Lehmann, 1986, pp. 48–52). Conditional expectations are not uniquely defined unless $\mathcal{Z}$ is a finite or countably infinite set and the probability that $\mathbf{Z} = \mathbf{z}$ is positive for all $\mathbf{z}$ in $\mathcal{Z}$; however, the lack of uniqueness in other cases has no practical effect. Define the conditional expectation of Y given $\mathbf{Z}$ so that the conditional expectation $E(Y| \mathbf{Z} = \mathbf{z})$ of Y given $\mathbf{Z} = \mathbf{z}$ is finite for all $\mathbf{z}$ in $\mathcal{Z}$. For $\mathbf{z}$ in $\mathcal{Z}$, let $G(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the convex real function on the real line with value $G(y, Y \mid \mathbf{Z} = \mathbf{z}; d) = E(d(Y - y)| \mathbf{Z} = \mathbf{z})$ at real y, and let $G\_(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the infimum of $G(Y| \mathbf{Z} = \mathbf{z})$. As in the case of $G(Y; d)$, the set $C(Y \mid \mathbf{Z} = \mathbf{z}; d)$ of real y such that $G(y, Y \mid \mathbf{Z} = \mathbf{z}; d) = G\_(Y \mid \mathbf{Z} = \mathbf{z}; d)$ is a closed and bounded nonempty interval $[c_L(Y \mid \mathbf{Z} = \mathbf{z}; d), c_U(Y \mid \mathbf{Z} = \mathbf{x}; d)]$ with midpoint $c(Y \mid \mathbf{Z} = \mathbf{z}; d)$. If d is strictly concave, $c(Y \mid \mathbf{Z} = \mathbf{z}; d) = c_L(Y \mid \mathbf{Z} = \mathbf{z}; d) = c_U(Y \mid \mathbf{Z} = \mathbf{z}; d)$ is the only member of $C(Y \mid \mathbf{Z}; d)$. Let $c(Y\|\mathbf{Z}; d)$ be $c(Y \mid \mathbf{Z} = \mathbf{z}; d)$ if $\mathbf{Z} = \mathbf{z}$ in T. Then $c(Y\|\mathbf{Z}; d)$ and $d(Y - c(Y\|\mathbf{Z}; d))$ are random variables. Let $G\_(Y\|\mathbf{Z}; d) = G\_(Y \mid \mathbf{Z} = \mathbf{z}; d)$ if $\mathbf{Z} = \mathbf{z}$ in R^K, so that $G\_(Y\|\mathbf{Z}; d) = d(Y - c(Y\|\mathbf{Z}; d))$. For any g in $L_{h(d)}$, $G\_(Y\|\mathbf{Z}; d) \le d(Y - g(\mathbf{Z}))$, so that $d(Y - c(Y\|\mathbf{Z}; d))$ has finite expectation $E(d(Y - c(Y\|\mathbf{Z}; d)) = E(G\_(Y\|\mathbf{Z}; d)) \le G(f, Y \mid \mathbf{Z}; d)$ for all f in $L_{h(d)}$, $c(Y\|\mathbf{Z}; d)$ is in $L_{h(d)}$, and $G\_(Y \mid \mathbf{Z}; d) = E(G\_(Y\|\mathbf{Z}; d))$.

In the case of quadratic discrepancy, $G-(Y; q) = \sigma^2(Y)$. For $\mathbf{z}$ in $\mathcal{Z}$, $c\left(Y \mid \mathbf{Z} = \mathbf{z}; q\right) = E(Y|\mathbf{Z} = \mathbf{z})$, $G\_\left(Y \mid \mathbf{Z} = \mathbf{z}; q\right)$ is the conditional variance $\sigma^2(Y| \mathbf{Z} = \mathbf{z}) = E([Y - E(Y| \mathbf{Z} = \mathbf{z})]^2 | \mathbf{Z} = \mathbf{z})$ of Y given $\mathbf{Z} = \mathbf{z}$, and $G\_(Y \mid \mathbf{Z}; d) = E(\sigma^2(Y\|\mathbf{Z}))$, where $\sigma^2(Y\|\mathbf{Z})$ is $\sigma^2(Y| \mathbf{Z} = \mathbf{z})$ if $\mathbf{Z} = \mathbf{z}$ in $\mathcal{Z}$ (Blackwell, 1947). In this case, if f and g are in $C(Y \mid \mathbf{Z}; q)$, then $g(\mathbf{Z}) = f(\mathbf{Z})$ with probability 1. In the case of absolute discrepancy, $G\_(Y; a)$ is the mean deviation $\mathrm{MD}(Y)$ of Y about the median, $G\_(Y| \mathbf{Z} = \mathbf{z})$ is the conditional mean deviation $\mathrm{MD}(Y| \mathbf{Z} = \mathbf{z})$ of Y given $\mathbf{Z} = \mathbf{z}$ for $\mathbf{z}$ in $\mathcal{Z}$, and $G\_(Y \mid \mathbf{Z}; a) = E(\mathrm{MD}(Y\|\mathbf{Z}))$, where $\mathrm{MD}(Y\|\mathbf{Z})$ is $\mathrm{MD}(Y \mid \mathbf{Z} = \mathbf{z}$ if $\mathbf{Z} = \mathbf{z}$ in $\mathcal{Z}$ (Haberman, 1996, pp. 396–398).

If Y and $\mathbf{Z}$ are independent, then the conditional distribution of Y given $\mathbf{Z}$ may be defined so that, for f in $L_{h(d)}(\mathbf{Z})$, the distribution of $d(Y - f(\mathbf{Z}))$ given $\mathbf{Z} = \mathbf{z}$ in $\mathcal{Z}$ is the same as the unconditional distribution of $d(Y - f(\mathbf{z}))$. It follows that $G\_(Y\|\mathbf{Z}; d)$ is the constant function $G\_(Y; d)$, $G\_(Y \mid \mathbf{Z}; d) = G\_(Y; d)$, and $\lambda(Y \mid \mathbf{Z}; d) = 0$. For any nonempty subset F of $L_{h(d)}(\mathbf{Z})$ that includes all constant functions on R^K, it then follows that $G\_(Y \mid \mathbf{Z}; F, d) = G\_(Y; d)$ and $\lambda(Y \mid \mathbf{Z}; F, d) = 0$. At the other extreme, if $\lambda(Y \mid \mathbf{Z}; d) = 1$, then f in $L_{h(d)}(\mathbf{Z})$ exists such that $Y = f(\mathbf{Z})$ with probability 1. If f is in F, $T = Y - f(\mathbf{Z})$ is independent of $\mathbf{Z}$, and $G\_(T; d) = E(d(T))$, then $G\_(T\|\mathbf{Z}; d)$ has the constant value $E(d(T))$, so that $G(f,$

Table 2 Probabilities $P(Y_2 = y, Z_1 = z_1)$

y	$P(Y = y, Z_1 = 1)$	$P(Y = y, Z_1 = 2)$	$P(Y = y, Z_1 = 3)$
1	.1	.2	.2
2	.2	.1	.2

$Y \mid \mathbf{Z}; d) = G\_(Y \mid \mathbf{Z}; d) = E(d(T))$, and f is in $C(Y \mid \mathbf{Z}; F, d)$. In the case of quadratic discrepancy, the condition on $G\_(T; d)$ is just the requirement that $E(T) = 0$. In the case of absolute discrepancy, the corresponding requirement is that T have a median of 0.

One common special case of interest arises if Z_k is in $L_{h(d)}$ for $1 \le k \le K$. Let A_K be the set of affine real functions on R^K, so that f is in A_K if and only if $f((1-c)\mathbf{z}_1 + c\mathbf{z}_2) = (1-c)f(\mathbf{z}_1) + cf(\mathbf{z}_2)$ if $\mathbf{z}_1$ and $\mathbf{z}_2$ are in R^K and c is real. If f is in A_K, then the subset $f^{-1}(0)$ of R^K such that $f(\mathbf{z}) = 0$ if and only if $\mathbf{z}$ is in $f^{-1}(0)$ is closed and convex (Rockafellar, 1970, pp. 6–10). Unless f is identically 0, $f^{-1}(0)$ has dimension less than K and an empty interior. Let $A(\mathbf{Z})$ be the set of real functions f on $\mathcal{Z}$ that are restrictions to $\mathcal{Z}$ of affine real functions on R^K. If f is in $A(\mathbf{Z})$ and $f(\mathbf{z})$ is not 0 for some $\mathbf{z}$ in $\mathcal{Z}$, then the probability that $f(\mathbf{Z}) = 0$ is less than 1, for otherwise a closed convex set with an empty interior would exist that includes $\mathbf{Z}$ with probability 1. If f is not identically 0, $d(Y - af(\mathbf{Z}))$ approaches ∞ with positive probability as $|a|$ approaches ∞, so that $E(d(Y - af(\mathbf{Z})))$ approaches ∞ if $|a|$ approaches ∞. It follows that $C(Y|\,\mathbf{Z}; A(\mathbf{Z}), d)$ is a nonempty and bounded convex set (Rockafellar, 1970, pp. 264–265). In the case of quadratic dispersion, $C(Y|\,\mathbf{Z}; A(\mathbf{Z}), q)$ has a single element $c(Y|\,\mathbf{Z}; A(\mathbf{Z}), d)$, and $G\_(Y|\,\mathbf{Z}; \mathcal{L}, q)$ is the infimum $\sigma^{(Y)}[1 - \rho^2(Y|\,\mathbf{Z})]$ of $\sigma^2(Y - \mathbf{b}'\mathbf{Z})$ for $\mathbf{b}$ in R^K, where the coefficient of determination $\rho^2(Y|\,\mathbf{Z})$ is the maximum squared correlation of Y and $\mathbf{b}'\mathbf{Z}$ for $\mathbf{b}$ in R^K with some nonzero element. It follows that $\lambda(Y|\,\mathbf{Z}; A(\mathbf{Z}), q) = \rho^2(Y|\,\mathbf{Z})$. If Y and $\mathbf{Z}$ have a joint multivariate normal distribution, then $\lambda(Y \mid \mathbf{Z}; q) = \lambda(Y|\,\mathbf{Z}; A(\mathbf{Z}), q)$, $G\_(Y;\ a) = (2/\pi)^{1/2}\sigma(Y)$, $G\_(Y|\,\mathbf{Z}; A(\mathbf{Z}), a) = G\_(Y \mid \mathbf{Z};\ a) = (2/\pi)^{1/2}\sigma(Y)[1 - \rho^2(Y|\,\mathbf{Z})]^{1/2}$, and $\lambda(Y \mid \mathbf{Z}; a) = \lambda(Y|\,\mathbf{Z}; A(\mathbf{Z}), a) = 1 - [1 - \rho(Y|\,\mathbf{Z})]^{1/2}$. Without the assumption of multivariate normality, $G\_(Y|\,\mathbf{Z}; A(\mathbf{Z}), a)$ is the infimum of $\mathrm{MD}(Y - \mathbf{b}'\mathbf{Z})$ for $\mathbf{b}$ in R^K, so that $G\_(Y \mid \mathbf{Z}; A(\mathbf{Z}, d)$ involves minimization of mean absolute error (Bloomfield & Steiger, 1983).

To illustrate lack of symmetry, consider $K = 1$ and $\mathbf{Y}$ is the random vector with the single element Y. Then $\lambda(Y \mid \mathbf{Z}; d)$ and $\lambda(Z_1 \mid \mathbf{Y}; d)$ may differ. Consider Table 2. Here $\lambda(Y \mid \mathbf{Z}; a) = .2$ and $\lambda(Z_1 \mid \mathbf{Y}; a) = 0$.

Nominal Prediction

Prediction of a nominal variable by probability prediction may also be considered. Let r be a positive integer, and let Y be in $\bar{r}$. For y in $\bar{r}$, let $\eta_r(y; d)$ be the nonnegative convex function on Π_r with value $\eta_r\left(\mathbf{p}, y; d\right) = \sum_{i=1}^{r} d\left(\delta_y(i) - p_i\right)$. Let $B_r(\mathbf{Z})$ be the convex set of functions $\mathbf{f}$ from $\mathcal{Z}$ to Π_r such that $\mathbf{f}(\mathbf{Z})$ is a random vector. Let F_r be a subset of $B_r(\mathbf{Z})$ that includes all constant functions on $\mathcal{Z}$ with value in Π_r. Let $G_r(Y; d)$ be the nonnegative convex function on Π_r with value $G_r(\mathbf{p}, Y; d) = E(\eta_r(\mathbf{p}, Y; d))$ at $\mathbf{p}$ in Π_r, and let $G_{r\text{-}}(Y;d)$ be the infimum of $G_r(Y; d)$. Because Π_r is closed and bounded, a nonempty closed convex set $C_r(Y; d)$ exists such that $\mathbf{p}$ in Π_r is in $C_r(Y; d)$ if and only if $G_r(\mathbf{p}, Y; d) = G_{r-}(Y; d)$. The set contains only one member $\mathbf{c}_r(Y; d)$ if d is strictly convex. Let $G_r(Y \mid \mathbf{Z}; d)$ be the real nonnegative convex function on $B_r(\mathbf{Z})$ with value $G_r(\mathbf{f}, Y \mid \mathbf{Z}; d) = E(\eta_r(\mathbf{f}(\mathbf{Z}), Y; d))$ at $\mathbf{f}$ in $B_r(\mathbf{Z})$. Let $G_{r-}(Y|\,\mathbf{Z}; F_r, d) \ge 0$ be the infimum of $G_r(Y \mid \mathbf{Z}, d)$ on F_r. Let $C_r(Y|\,\mathbf{Z}; F_r, d)$ be the set of $\mathbf{f}$ in F_r such that $\mathbf{f}$ is in $C_r(Y \mid \mathbf{Z}; d)$ if and only if $G_r(\mathbf{f}, Y \mid \mathbf{Z}; d) = G_{r-}(YI\mathbf{Z}; F_r, d)$. If F_r is convex and $C_r(Y \mid \mathbf{Z}; d)$ is nonempty, then $C_r(Y|\,\mathbf{Z}; F_r, d)$ is convex. If $\mathbf{f}$ in F_r exists such that $\boldsymbol{\delta}_r(Y) = \mathbf{f}(\mathbf{Z})$ with probability 1, then $G_{r-}(Y|\,\mathbf{Z}; F_r, d) = 0$. The dispersion measure $G_{r\,-}\ (Y; d)$ is 0 if and only if y in $\bar{r}$ exists such that $Y = y$ with probability 1. If no y in $\bar{r}$ exists such that $Y = y$ with probability 1, then the measure of proportional reduction in error for F_r and d is

$$\lambda_r\left(Y|\mathbf{Z}; F_r, d\right) = 1 - \frac{G_{r-}\left(Y|\mathbf{Z}; F_r, d\right)}{G_{r-}\left(Y; d\right)}.$$

The value of $\lambda_r(Y|\,\mathbf{Z}; F_r, d)$ is nonnegative. If $\mathbf{f}$ in F_r exists such that $\boldsymbol{\delta}_r(Y) = \mathbf{f}(\mathbf{Z})$ with probability 1, then $\lambda_r(Y|\,\mathbf{Z}; F_r, d) = 1$. Let $G_{r-}(Y \mid \mathbf{Z}; d) = G_{r-}(Y|\,\mathbf{Z}; B_r(\mathbf{Z}), d)$ and $\lambda_r(Y \mid \mathbf{Z}; d) = \lambda_r(Y|\,\mathbf{Z}; B_r(\mathbf{Z}), d)$. Then $G_{r-}(Y|\,\mathbf{Z}; F_r, d) \ge G_{r-}(Y \mid \mathbf{Z}; d)$ and $\lambda_r(Y|\,\mathbf{Z}; F_r, d) \le \lambda_r(Y \mid \mathbf{Z}; d)$. The set $C_r(Y \mid \mathbf{Z}; d) = C_r(Y|\,\mathbf{Z}; B_r(\mathbf{Z}), d)$.

The conditional distribution of Y given $\mathbf{Z}$ can be used to show that $C_r(Y \mid \mathbf{Z}; d)$ is not empty and therefore convex. For $\mathbf{z}$ in $\mathcal{Z}$, let $G_{r-}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the infimum of $G_r(Y \mid \mathbf{Z} = \mathbf{z}; d)$, where $G_r(Y \mid \mathbf{Z} = \mathbf{z}; d)$ is the function on Π_r with value

$G_r(p, Y \mid \mathbf{Z}=\mathbf{z}; d)$ at $\mathbf{p}$ in Π_r equal to the conditional expectation of $\eta_r(\mathbf{p}, Y; d)$ given $\mathbf{Z}=\mathbf{z}$. Then the set $C_r(Y \mid \mathbf{Z}=\mathbf{z}; d)$ of $\mathbf{p}$ in Π_r that satisfy $G_r(\mathbf{p}, Y \mid \mathbf{Z}=\mathbf{z}; d)=G_{r-}(Y \mid \mathbf{Z}=\mathbf{z}; d)$ is nonempty, bounded, closed, and convex. The set has a single member $\mathbf{c}_r(Y \mid \mathbf{Z}=\mathbf{z}; d)$ if d is strictly convex. Let $G_{r-}(Y \| \mathbf{Z}; d)=G_{r-}(Y \mid \mathbf{Z}=\mathbf{z}; d)$ if $\mathbf{Z}=\mathbf{z}$ in $\mathcal{Z}$. Then $G_{r-}(Y \mid \mathbf{Z}; d)=E(G_{r-}Y \| \mathbf{Z}; d))$. There exists $\mathbf{f}$ in $B_r(\mathbf{Z})$ such that $\mathbf{f}(\mathbf{z})$ is in $C_r(Y \mid \mathbf{Z}=\mathbf{z}; d)$ for all $\mathbf{z}$ in $\mathcal{Z}$ (Brown & Purves, 1973), so that $G_{r-}(Y \mid \mathbf{Z}; d)=G_r(\mathbf{f}, Y \mid \mathbf{Z}; d)$. Thus $\mathbf{f}$ is in $C_r(Y \mid \mathbf{Z}; d)$. It follows that the error measure $G_{r-}(Y \mid \mathbf{Z}; d)=0$ if and only if $\mathbf{f}$ in $B_r(\mathbf{Z})$ exists such that $Y=\mathbf{f}(\mathbf{Z})$ with probability 1. As in the case of prediction of a numerical variable, it follows that $G_{r-}(Y \mid \mathbf{Z}; F_r, d)=G_{r-}(Y; d)$ and $\lambda_r(Y \mid \mathbf{Z}; F_r, d)=0$ if Y and $\mathbf{Z}$ are independent.

The simplest cases of F_r involve a function $\mathbf{h}$ on $\mathcal{Z}$ onto an H-dimensional finite set $\mathcal{H}$ such that $\mathbf{h}(\mathbf{Z})$ is a random vector and $P(\mathbf{h}(\mathbf{Z})=\mathbf{u})>0$ for each $\mathbf{u}$ in $\mathcal{H}$. The set $B_r(\mathbf{h}, \mathbf{Z})$ consists of all functions $\mathbf{f}=\mathbf{g}(\mathbf{h})$ such that $\mathbf{g}$ is a function from $\mathcal{H}$ to Π_r. In this case, the conditional distribution of Y given $\mathbf{h}(\mathbf{Z})$ is uniquely defined, $G_{r-}(Y \mid \mathbf{Z}; B_r(\mathbf{h}, \mathbf{Z}), d)=G_{r-}(Y \mid \mathbf{h}(\mathbf{Z}); d)$, and $\lambda_r(Y \mid \mathbf{Z}; B_r(\mathbf{h}, \mathbf{Z}), d)=\lambda_r(Y \mid \mathbf{h}(\mathbf{Z}); d)$.

For nominal predicted variables, quadratic discrepancy is quite attractive, for $G_{r-}\left(Y; q\right) = 1 - \sum_{y=1}^{r} \left[P\left(Y = y\right)\right]^2$ is the coefficient of concentration of Y (Haberman, 1982a). If Y and Y' are independent random variables with the same distribution, then $G_{r-}(Y; q)=P(Y \neq Y')$. If Y and Y'' are conditionally independent given $\mathbf{Z}$ and have the same conditional distribution given $\mathbf{Z}$, then $G_{r-}(Y \mid \mathbf{Z}; q)=P(Y \neq Y'')$ is the expected conditional coefficient of concentration of Y given $\mathbf{Z}$. The measure $\lambda_r(Y \mid \mathbf{Z}; q)$ corresponds to the tau measure (Goodman & Kruskal, 1954; Haberman, 1982a, 1982b). Here $\lambda_r(Y \mid \mathbf{Z}; q)=0$ if and only if Y and $\mathbf{Z}$ are independent. The case of absolute discrepancy also leads to a simple result. Here $G_{r-}(Y; a)=1-\max_{1 \le y \le r} P(Y=y)$, and $G_{r-}(Y \mid \mathbf{Z}; a)$ is the expectation of $1-\max_{1 \le y \le r} P(Y=y \| \mathbf{Z})$, where $P(Y=y \mid \mathbf{Z})$ has value $P(Y=y \mid \mathbf{Z}=\mathbf{z})$ if $\mathbf{Z}=\mathbf{z}$ in $\mathcal{Z}$ and $P(Y=y \mid \mathbf{Z}=\mathbf{z})$ is the conditional probability that $Y=y$ given $\mathbf{Z}=\mathbf{z}$. The measure $\lambda_r(Y \mid \mathbf{Z}; a)$ corresponds to the lambda measure of Goodman and Kruskal. For $r=2$, $\lambda_r(Y \mid \mathbf{Z}; a)=\lambda(Y \mid \mathbf{Z}; a)$. As in the case of prediction of a numeric variable, the disadvantage of absolute discrepancy is that $\lambda_r(Y \mid \mathbf{Z}; a)$ can be 0 even if Y and $\mathbf{Z}$ are not independent.

To illustrate the problem with $\lambda_r(Y \mid \mathbf{Z}; a)$, consider Table 2 for $K=1$. Here $\lambda_2(Y \mid \mathbf{Z}; a)=0.2$ but $\lambda_3(Z_1 \mid \mathbf{Y}; a)=0$. The difference in the number of categories for Y and Z_1 is not essential here. If Z_1 only had values 1 and 2, $P(Y=Z_1=1)=P(Y=1, Z_1=2)=1/4$, $P(Y=2, Z_1=1)=1/12$, and $P(Y=Z_1=2)=5/12$, then $\lambda_2(Y \mid \mathbf{Z}; a)=1/3$ and $\lambda_2(Z_1 \mid \mathbf{Y})=0$.

Estimation of Measures

Estimation of measures of agreement and prediction accuracy varies somewhat in difficulty. To begin, measures of absolute agreement are considered. These measures are relatively straightforward to estimate. The kappa family is then examined. Here results are straightforward in a sense, but implementation must be considered. Lambda measures for agreement add complications related to the minimization portion of the definition. Measures of prediction involve issues quite similar to those for lambda measures for agreement and also can involve further problems associated with spaces of excessive dimension. For discussion of many cases in which all variables under study are nominal, see Goodman and Kruskal (1963, 1972).

Measures of Absolute Agreement

Let X_j be in $L_{h(d)}$ for $1 \le j \le J$ and let no real x exist such that $X_j = x$ with probability 1 for $1 \le j \le J$. Then the initial $n \ge 1$ observations $\mathbf{X}_i$, $1 \le i \le n$, yield an unbiased estimate $\hat{D}_{Jn}(\mathbf{X}; d) = n^{-1} \sum_{i=1}^{n} d_J\left(\mathbf{X}_i\right)$ of $D_J(\mathbf{X}; d)$ that, by the strong law of large numbers, converges with probability 1 to $D_J(\mathbf{X}; d)$ as n approaches ∞.

In the nominal case, all elements of $\mathbf{X}$ are positive integers no greater than the positive integer $r \ge 2$. Then the unbiased estimate $\hat{D}_{Jrn}(\mathbf{X}) = n^{-1} \sum_{i=1}^{n} d_{Jr}\left(\mathbf{X}_i\right)$ of $D_{Jr}(\mathbf{X}; d)$ converges with probability 1 to $D_{Jr}(\mathbf{X})$ as n approaches ∞.

Under the added condition that X_j is in $L_{2h(d)}$ for $1 \le j \le J$, the variance $\sigma^2\left(\hat{D}_{Jn}(\mathbf{X}; d)\right)$ is $n^{-1}\sigma^2(d_J(\mathbf{X}))$. As the sample size n approaches ∞,

$$n^{1/2}\left[\hat{D}_{Jn}(\mathbf{X}; d) - D_J(\mathbf{X}; d)\right]$$

converges in distribution to a normal random variable with mean 0 and variance $\sigma^2(d_J(\mathbf{X}))$.

In addition, in the nominal case in which X_j is in $\bar{r}$ for $1 \le j \le J$, $\sigma^2\left(\hat{D}_{Jrn}(\mathrm{X})\right) = n^{-1}\sigma^2\left(d_{Jr}(\mathrm{X})\right)$. As n approaches ∞,

$$n^{1/2}\left[\hat{D}_{Jrn}(\mathrm{X}) - D_{Jr}(\mathrm{X})\right]$$

converges in distribution to a normal random variable with mean 0 and variance $\sigma^2(d_{Jr}(\mathrm{X}))$.

For $n \ge 2$, the estimate

$$s_n^2\left(d_J(\mathrm{X})\right) = (n-1)^{-1}\sum_{i=1}^{n}\left[d_J\left(\mathrm{X}_i\right) - \hat{D}_{Jn}(\mathrm{X};d)\right]^2$$

is an unbiased estimate of $\sigma^2(d_J(\mathrm{X}))$, and $s_n^2\left(d_J(\mathrm{X})\right)$ converges to $\sigma^2(d_J(\mathrm{X}))$ with probability 1 as n approaches ∞. Thus $s_n^2\left(\hat{D}_{Jn}(\mathrm{X};d)\right) = n^{-1}s_n^2\left(d_J(\mathrm{X})\right)$ provides an unbiased estimate of $\sigma^2\left(\hat{D}_{Jn}(\mathrm{X};d)\right)$. An asymptotic confidence interval for $D_J(\mathrm{X};d)$ is available. Let $s_n\left(\hat{D}_{Jn}(\mathrm{X};d)\right)$ be the nonnegative square root of $s_n^2\left(\hat{D}_{Jn}(\mathrm{X};d)\right)$. Let $0 < \alpha < 1$, and let $z_{\alpha/2}$ be the real number such that a standard normal random variable exceeds $z_{\alpha/2}$ with probability $\alpha/2$. If $d_J(\mathrm{X})$ does not equal $D_J(\mathrm{X};d)$ with probability 1, then, as n approaches ∞, the probability approaches $1-\alpha$ that

$$\mid D_J(\mathrm{X};d) - \hat{D}_{Jn}(\mathrm{X};d) \mid \le z_{\alpha/2}s_n\left(\hat{D}_{Jn}(\mathrm{X};d)\right).$$

The same argument for the nominal case with X_j in $\bar{r}$ for $1 \le j \le J$ shows that

$$s_n^2\left(d_{Jr}(\mathrm{X})\right) = (n-1)^{-1}\sum_{i=1}^{n}\left[d_{Jr}\left(\mathrm{X}_i\right) - \hat{D}_{Jrn}(\mathrm{X})\right]^2$$

is an unbiased estimate of $\sigma^2(d_{Jr}(\mathrm{X}))$, and $s_n^2\left(d_{Jr}(\mathrm{X})\right)$ converges to $\sigma^2(d_{Jr}(\mathrm{X}))$ with probability 1 as n approaches ∞. Thus $s_n^2\left(\hat{D}_{Jrn}(\mathrm{X})\right) = n^{-1}s_n^2\left(d_{Jr}(\mathrm{X})\right)$ provides an unbiased estimate of $\sigma^2\left(\hat{D}_{Jnr}(\mathrm{X})\right)$. For an asymptotic confidence interval, let $s_n\left(\hat{D}_{Jnr}(\mathrm{X})\right)$ be the nonnegative square root of $s_n^2\left(\hat{D}_{Jrn}(\mathrm{X})\right)$. If $d_{Jr}(\mathrm{X})$ does not equal $D_{Jr}(\mathrm{X})$ with probability 1, then, as n approaches ∞, the probability approaches $1-\alpha$ that

$$\mid D_{Jr}(\mathrm{X}) - \hat{D}_{Jrn}(\mathrm{X}) \mid \le z_{\alpha/2}s_n\left(\hat{D}_{Jrn}(\mathrm{X})\right).$$

Kappa Measures

Examination of kappa measures entails use of U-statistics (Hoeffding, 1948). Assume that $n \ge 2$. As in the case of absolute measures of agreement, assume that X_j is in $L_{h(d)}$ for $1 \le j \le J$ and no real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$. The U-statistic $\hat{D}_{Jn}\left(\mathrm{X_J};d\right)$ is the average of $e_J(\mathrm{X}_{i(1)},\mathrm{X}_{i(2)};d)$ for positive integers $i(1)$ and $i(2)$ such that $1 \le i(2) < i(1) \le n$. As n approaches ∞, $\hat{D}_{Jn}\left(\mathrm{X_J};d\right)$ converges to $D_J(\mathrm{X_J};d)$ with probability 1 (Hoeffding, 1948, 1961; Lee, 1990), and

$$\hat{\kappa}_{Jn}(\mathrm{X};d) = \frac{\hat{D}_{Jn}\left(\mathrm{X_J};d\right) - \hat{D}_{Jn}(\mathrm{X};d)}{\hat{D}_{Jn}\left(\mathrm{X_J};d\right)}$$

converges to $\kappa_J(\mathrm{X};d)$ with probability 1. The convention is used throughout the report that $0/0 = 0$. It should be noted that $\hat{D}_{Jn}\left(\mathrm{X_J};d\right) = 0$ implies that $\hat{D}_{Jn}(\mathrm{X};d)$ is also 0, so that $\hat{\kappa}_{Jn}(\mathrm{X};d) = 0$ if, for some real x, $X_{ij} = 0$ for $1 \le i \le n$ and $1 \le j \le J$.

In the nominal case, for $\mathbf{x}_1$ and $\mathbf{x}_2$ J-dimensional vectors with elements in $\bar{r}$, let $e_{Jr}(\mathbf{x}_1,\mathbf{x}_2)$ be $\sum_{y=1}^{r} e_J\left(\boldsymbol{\delta}_{Jy}\left(\mathbf{x}_1\right),\boldsymbol{\delta}_{Jy}\left(\mathbf{x}_2\right)\right)$. The choice of d has no effect, so the discrepancy function does not appear for $e_{Jr}(\mathbf{x}_1,\mathbf{x}_2)$. Then $\hat{D}_{Jnr}\left(\mathrm{X_J}\right)$ is the average of $e_{Jr}(\mathrm{X}_{i(1)},\mathrm{X}_{i(2)})$ for positive integers $i(1)$ and $i(2)$ such that $i(2) < i(1) \le n$. As n approaches ∞, $\hat{D}_{Jrn}\left(\mathrm{X_J}\right)$ converges to $D_{Jr}(\mathrm{X_J})$ with probability 1, and

$$\hat{\kappa}_{Jrn}(\mathrm{X}) = \frac{\hat{D}_{Jrn}\left(\mathrm{X_J}\right) - \hat{D}_{Jrn}(\mathrm{X})}{\hat{D}_{Jrn}\left(\mathrm{X_J}\right)}$$

converges to $\kappa_{Jr}(\mathrm{X})$ with probability 1.

Normal approximations are somewhat more complex. Let X_j be in $L_{2h(d)}$ for $1 \le j \le J$. For $\mathbf{x}$ in R^K, let $E_J(\mathbf{X};\ d)$ be the random variable with value $E(e_J(\mathbf{x},\ \mathbf{X};\ d))$ at a J-dimensional vector $\mathbf{x}$ if $\mathbf{X}=\mathbf{x}$. By Fubini's theorem, $E(E_J(\mathbf{X};\ d))=E(e_J(e_J(\mathbf{X}_1,\mathbf{X}_2;d))=D_J(\mathbf{X}_\mathbf{J};d)$. As the sample size n increases,

$$n^{1/2}\left[\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)-D_J\left(\mathbf{X}_\mathbf{J};d\right)\right]$$

converges in law to a normal random variable with mean 0 and variance $4\sigma^2(E_J(\mathbf{X};\ d))$. The variance $\sigma^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ satisfies the condition that $n\sigma^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ converges to $4\sigma^2(E_J(\mathbf{X};\ d))$, but an exact expression is a bit more complex. For $n \ge 3$,

$$\sigma^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)=\frac{1}{n(n-1)}\left[4(n-2)\sigma^2\left(E_J(\mathbf{X};d)\right)+2\sigma^2\left(e_J\left(\mathbf{X}_1,\mathbf{X}_2;d\right)\right)\right].$$

For $1 \le i \le n$, let $\hat{E}_{Jin}(\mathbf{X};d)$ be the average of $e_J(\mathbf{X}_i,\mathbf{X}_k;d)$ for positive integers $k \le n$ not equal to i, and let $s_n^2\left(E_J(\mathbf{X};d)\right)$ be the average of $\left[\hat{E}_{Jin}(\mathbf{X};d)-\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right]^2$ for $1 \le i \le n$. Let $s_n^2\left(e_J\left(\mathbf{X}_1,\mathbf{X}_2;d\right)\right)$ be the average of $\left[e_J\left(\mathbf{X}_i,\mathbf{X}_k;d\right)-\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right]^2$ for $1 \le k < i \le n$. For $n \ge 4$, an unbiased estimate of $\sigma^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ is

$$s_n^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)=[(n-2)(n-3)]^{-1}\left[4(n-1)s_n^2\left(E_J(\mathbf{X};d)\right)-2s_n^2\left(e_J\left(\mathbf{X}_1,\mathbf{X}_2;d\right)\right)\right].$$

Let $s_{na}^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)=4s_n^2\left(E_J(\mathbf{X};d)\right)/n$. Let $s_n\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ be the positive square root of $s_n^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ if $s_n^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)>0$, and let $s_n\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ be 0 otherwise. Let $s_{na}(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)$ be the nonnegative square root of $s_{na}^2(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)$. As the sample size n goes to ∞, $ns_n^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ and $ns_{na}^2\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right)$ both converge with probability 1 to $4\sigma^2(E_J(\mathbf{X};\ d))$. If $E_J(\mathbf{X};\ d)$ is not equal to $D_J(\mathbf{X}_\mathbf{J};d)$ with probability 1, then, as n approaches ∞, the probability approaches $1-\alpha$ that

$$\mid D_J\left(\mathbf{X}_\mathbf{J};d\right)-\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\mid\le z_{\alpha/2}s_n\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right),$$

and the probability approaches $1-\alpha$ that

$$\mid D_J\left(\mathbf{X}_\mathbf{J};d\right)-\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\mid\le z_{\alpha/2}s_{na}\left(\hat{D}_{Jn}\left(\mathbf{X}_\mathbf{J};d\right)\right).$$

In the nominal case, let all elements of $\mathbf{X}$ be positive integers no greater than r, and let $E_{Jr}(\mathbf{X})$ be the random variable with value $E(e_{Jr}(\mathbf{x},\mathbf{X}))$ at a J-dimensional vector $\mathbf{x}$ if $\mathbf{X}=\mathbf{x}$ and all elements of $\mathbf{x}$ are positive integers no greater than r. Then $n^{1/2}\left[\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)-D_{Jr}\left(\mathbf{X}_\mathbf{J}\right)\right]$ converges in law to a normal random variable with mean 0 and variance $4\sigma^2(E_{Jr}(\mathbf{X}))$. Here $n\sigma^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ converges to $4\sigma^2(E_{Jr}(\mathbf{X}))$, and for $n \ge 3$,

$$\sigma^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)=\frac{1}{n(n-1)}\left[4(n-2)\sigma^2\left(E_{Jr}(\mathbf{X})\right)+2\sigma^2\left(e_{Jr}\left(\mathbf{X}_1,\mathbf{X}_2\right)\right)\right].$$

For $1 \le i \le n$, let $\hat{E}_{Jrin}(\mathbf{X})$ be the average of $e_{Jr}(\mathbf{X}_i,\mathbf{X}_k)$ for positive integers $k \le n$ not equal to i, and let $s_n^2\left(E_{Jr}(\mathbf{X})\right)$ be the average of $\left[\hat{E}_{Jrin}(\mathbf{X})-\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right]^2$ for $1 \le i \le n$. Let $s_n^2\left(e_{Jr}\left(\mathbf{X}_1,\mathbf{X}_2\right)\right)$ be the average of $\left[e_{Jr}\left(\mathbf{X}_i,\mathbf{X}_k\right)-\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right]^2$ for $1 \le k < i \le n$. For $n \ge 4$, an unbiased estimate of $\sigma^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ is

$$s_n^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J};d\right)\right)=[(n-2)(n-3)]^{-1}\left[4(n-1)s_n^2\left(E_{Jr}(\mathbf{X})\right)-2s_n^2\left(e_{Jr}\left(\mathbf{X}_1,\mathbf{X}_2\right)\right)\right].$$

Let $s_{na}^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)=4s_n^2\left(E_{Jr}(\mathbf{X})\right)/n$. Let $s_n\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ be the positive square root of $s_n^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ if $s_n^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)>0$, and let $s_n\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ be 0 otherwise. Let $s_{na}(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)$ be the nonnegative square root of $s_{na}^2(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)$. As the sample size n goes to ∞, $ns_n^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ and $ns_{na}^2\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right)$ both converge with probability 1 to $4\sigma^2(E_{Jr}(\mathbf{X}))$. If $E_{Jr}(\mathbf{X})$ is not equal to $D_{Jr}(\mathbf{X})$ with probability 1, then, as n approaches ∞, the probability approaches $1-\alpha$ that

$$\mid D_{Jr}\left(\mathbf{X}_\mathbf{J}\right)-\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\mid\le z_{\alpha/2}s_n\left(\hat{D}_{Jrn}\left(\mathbf{X}_\mathbf{J}\right)\right).$$

The probability also approaches 1 − α that

$$| D_{Jr}\left(\mathrm{X_J}\right) - \widehat{D}_{Jrn}\left(\mathrm{X_J}\right) | \le z_{\alpha/2} s_{na}\left(\widehat{D}_{Jrn}\left(\mathrm{X_J}\right)\right).$$

The normal approximation for $\widehat{\kappa}_{Jn}(\mathbf{X}; d)$ is a bit more complicated because a linearization step is needed and $\widehat{\kappa}_{Jn}(\mathbf{X}; d)$ is not unbiased. For J-dimensional vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, let

$$e_{J\kappa}\left(\mathbf{x}_1, \mathbf{x}_2; d\right) = \left[D_{JJ}\left(\mathrm{X_J}; d\right)\right]^{-1}\left\{\frac{1}{2}[d_J\left(\mathbf{x}_1\right) + d_J(\mathbf{x}_2] - \kappa_J(\mathbf{X}; d)\, e_J(\mathbf{x}_1, \mathbf{x}_2); d)\right\}.$$

Let $E_{J\kappa}(\mathbf{X}; d)$ be the random variable with value $E(e_{J\kappa}(\mathbf{x}, \mathbf{X}; d))$ at a J-dimensional vector $\mathbf{x}$ if $\mathbf{X} = \mathbf{x}$. As n approaches ∞, $n^{1/2}\left[\widehat{\kappa}_{Jn}(\mathbf{X}; d) - \kappa_J(\mathbf{X}; d)\right]$ converges in law to a normal random variable with mean 0 and variance $4\sigma^2(E_{J\kappa}(\mathbf{X}; d))$. The asymptotic variance $\sigma_a^2\left(\widehat{\kappa}_{Jn}\left(\mathrm{X_J}; d\right)\right) = 4\sigma^2\left(E_{J\kappa}(\mathbf{X}; d)\right)/n$. No simple formula is available for the variance of $\widehat{\kappa}_{Jn}(\mathbf{X}; d)$.

For an asymptotic confidence interval, let

$$\widehat{e}_{J\kappa}\left(\mathbf{X}_i, \mathbf{X}_k; d\right) = \left[\widehat{D}_{Jn}\left(\mathrm{X_J}; d\right)\right]^{-1}\left\{\frac{1}{2}[d_J\left(\mathbf{X}_i\right) + d_J(\mathbf{X}_k] - \widehat{\kappa}_{Jn}(\mathbf{X})\, e_J(\mathbf{X}_i, \mathbf{X}_k))\right\}$$

for distinct positive integers i and k no greater than n. For $1 \le i \le n$, let $\widehat{E}_{J\kappa in}(\mathbf{X})$ be the average of $\widehat{e}_{J\kappa r}\left(\mathbf{X}_i, \mathbf{X}_k\right)$ for positive integers $k \le n$ not equal to i. The average of the $\widehat{E}_{J\kappa in}(\mathbf{X}; d)$, $1 \le i \le n$, is 0. Let $s_{na}^2\left(\widehat{\kappa}_{Jn}(\mathbf{X}; d)\right)$ be $4n^{-2}$ times the sum of $\left[\widehat{E}_{J\kappa in}(\mathbf{X}; d)\right]^2$ for $1 \le i \le n$. As the sample size n goes to ∞, $ns_{na}^2\left(\widehat{\kappa}_{Jn}(\mathbf{X}; d)\right)$ converges with probability 1 to $4\sigma^2(E_{J\kappa}(\mathbf{X}; d))$. If $E_{J\kappa}(\mathbf{X}; d))$ is not 0 with probability 1, then the probability approaches 1 − α that

$$| \kappa_J(\mathbf{X}; d) - \widehat{\kappa}_{Jn}(\mathbf{X}; d) | \le z_{\alpha/2} s_{na}\left(\widehat{\kappa}_{Jn}(\mathbf{X}; d)\right).$$

In the nominal case, let all elements of $\mathbf{X}$ be positive integers no greater than r. Let

$$e_{J\kappa r}\left(\mathbf{x}_1, \mathbf{x}_2\right) = \left[D_{Jr}\left(\mathrm{X_J}; d\right)\right]^{-1}\left\{\frac{1}{2}\left[d_{Jr}\left(\mathbf{x}_1\right) + d_{Jr}\left(\mathbf{x}_2\right)\right] - \kappa_{Jr}(\mathbf{X})\, e_{Jr}(\mathbf{x}_1, \mathbf{x}_2))\right\}$$

for J-dimensional vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ with all elements positive integers no greater than r. Let $E_{J\kappa r}(\mathbf{X})$ be the random variable with value $E(e_{J\kappa r}(\mathbf{x}, \mathbf{X}))$ at a J-dimensional vector $\mathbf{x}$ if $\mathbf{X} = \mathbf{x}$. As n approaches ∞, $n^{1/2}\left[\widehat{\kappa}_{Jrn}(\mathbf{X}) - \kappa_{Jr}(\mathbf{X})\right]$ converges in law to a normal random variable with mean 0 and variance $4\sigma^2(E_{J\kappa r}(\mathbf{X}))$. The asymptotic variance $\sigma_{na}^2\left(\widehat{\kappa}_{Jrn}\left(\mathrm{X_J}\right)\right) = 4\sigma^2\left(E_{J\kappa r}(\mathbf{X})\right)/n$.

For distinct positive integers i and k no greater than n, let

$$\widehat{e}_{J\kappa rn}\left(\mathbf{X}_i, \mathbf{X}_k; d\right) = \left[\widehat{D}_{Jrn}\left(\mathrm{X_J}; d\right)\right]^{-1}\left\{\frac{1}{2}\left[d_{Jr}\left(\mathbf{X}_i\right) + d_{Jr}\left(\mathbf{X}_k\right)\right] - \widehat{\kappa}_{Jrn}(\mathbf{X})\, e_{Jr}(\mathbf{X}_i, \mathbf{X}_k))\right\}.$$

For $1 \le i \le n$, let $\widehat{E}_{J\kappa rin}(\mathbf{X})$ be the average of $\widehat{e}_{J\kappa r}\left(\mathbf{X}_i, \mathbf{X}_k\right)$ for positive integers $k \le n$ not equal to i. Let $s_{na}^2\left(\widehat{\kappa}_{Jrn}(\mathbf{X})\right)$ be $4n^{-2}$ times the sum of $\left[\widehat{E}_{J\kappa rin}(\mathbf{X})\right]^2$ for $1 \le i \le n$. As the sample size n goes to ∞, $ns_{na}^2\left(\widehat{\kappa}_{Jrn}(\mathbf{X})\right)$ converges with probability 1 to $4\sigma^2(E_{J\kappa r}(\mathbf{X}))$. If $E_{J\kappa r}(\mathbf{X})$ does not equal 0 with probability 1, then the probability approaches 1 − α that

$$| \kappa_{Jr}(\mathbf{X}) - \widehat{\kappa}_{Jrn}(\mathbf{X}) | \le z_{\alpha/2} s_{na}\left(\widehat{\kappa}_{Jrn}(\mathbf{X})\right).$$

Because, except for the nominal case, no assumption has been made that the variables X_j, $1 \le j \le J$, are discrete, these results generalize very similar results for kappa estimates for the case of all X_j confined to the integers 1 to r (Fleiss et al., 1969).

Lambda Measures of Agreement

In the case of the lambda measures for agreement, let the X_j, $1 \le j \le J$, be in $L_{h(d)}$, and assume that no real x exists such that $X_j = x$ with probability 1 for $1 \le j \le J$. For $n \ge 1$, let $\widehat{G}_{Jn}(\mathbf{X}; d)$ be the function on the real line with value $\widehat{G}_{Jn}(x, \mathbf{X}; d) = n^{-1}\sum_{i=1}^{n}\eta_J\left(x, \mathbf{X}_i; d\right)$ for real x. Let $\widehat{G}_{Jn-}(\mathbf{X}; d)$ be the infimum of $\widehat{G}_{Jn}(\mathbf{X}; d)$. Let $\widehat{\lambda}_{Jan}(\mathbf{X}; d) = \left[\widehat{G}_{Jn-}(\mathbf{X}; d) - \widehat{D}_J(\mathbf{X}; d)\right]/\widehat{G}_{Jn-}(\mathbf{X}; d)$. As n approaches ∞, $\widehat{G}_{Jn-}(\mathbf{X}; d)$ converges to $G_{J-}(\mathbf{X}; d)$ with probability 1, and $\widehat{\lambda}_{Jan}(\mathbf{X}; d)$ converges to $\lambda_{Ja}(\mathbf{X}; d)$ with probability 1 (Haberman, 1989).

In the nominal case, for $n \geq 1$, let $\widehat{G}_{Jrn}(\mathbf{X}; d)$ be the function on Π_r with value $\widehat{G}_{Jrn}\left(\mathbf{p}, \mathbf{X}; d\right) = n^{-1} \sum_{i=1}^{n} \eta_{Jr}\left(\mathbf{p}, \mathbf{X}_i; d\right)$ for $\mathbf{p}$ in Π_r. Let $\widehat{G}_{Jrn-}(\mathbf{X}; d)$ be the infimum of $\widehat{G}_{Jrn}(\mathbf{X}; d)$. Let $\widehat{\lambda}_{Jarn}(\mathbf{X}; d) = \left[\widehat{G}_{Jrn-}(\mathbf{X}; d) - \widehat{D}_{Jr}(\mathbf{X})\right] / \widehat{G}_{Jrn-}(\mathbf{X}; d)$. As n approaches ∞, $\widehat{G}_{Jrn-}(\mathbf{X}; d)$ converges to $G_{Jr-}(\mathbf{X}; d)$ with probability 1, and $\widehat{\lambda}_{Jarn}(\mathbf{X}; d)$ converges to $\lambda_{Jar}(\mathbf{X}; d)$ with probability 1.

For a normal approximation, add the conditions that X_j be in $L_{2h(d)}$ for $1 \leq i \leq n$ and that the set $C_J(\mathbf{X}; d)$ of minima of $G_J(\mathbf{X}; d)$ has a single element $c_J(\mathbf{X}; d)$, as is always the case if $d = q$. As n approaches ∞, $n^{1/2}\left[\widehat{G}_{J-}(\mathbf{X}; d) - G_{J-}(\mathbf{X}; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_J(c_J(\mathbf{X}; d), \mathbf{X}; d))$. Let

$$\eta_{J\lambda}(\mathbf{X}; d) = \left[G_{J-}(\mathbf{X}; d)\right]^{-1}\left[d_J(\mathbf{X}) - \lambda_{Ja}(\mathbf{X}; d)\, \eta_J\left(c_J(\mathbf{X}; d), \mathbf{X}; d\right)\right].$$

As n approaches ∞, $n^{1/2}\left[\widehat{\lambda}_{Jan}(\mathbf{X}; d) - \lambda_{Ja}(\mathbf{X}; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_{J\lambda}(\mathbf{X}; d))$.

In the nominal case, let $C_{Jr}(\mathbf{X}; d)$ have a single element $\mathbf{c}_{Jr}(\mathbf{X}; d)$, as is always the case if $d = q$. This situation also holds for $d = a$ if a y in $\overline{r}$ exists such that $p_{Jy}(\mathbf{X}) > p_{Jy'}(\mathbf{X})$ for all y' in $\overline{r}$ not equal to y. As n approaches ∞, $n^{1/2}\left[\widehat{G}_{Jr-}(\mathbf{X}; d) - G_{Jr-}(\mathbf{X}; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_{Jr}(\mathbf{c}_{Jr}(\mathbf{X}; d), \mathbf{X}; d))$. Let

$$\eta_{J\lambda r}(\mathbf{X}; d) = \left[G_{Jr-}(\mathbf{X}; d)\right]^{-1}\left[d_{Jr}(\mathbf{X}) - \lambda_{Jar}(\mathbf{X}; d)\, \eta_{Jr}\left(\mathbf{c}_{Jr}(\mathbf{X}; d), \mathbf{X}; d\right)\right].$$

As n approaches ∞, $n^{1/2}\left[\widehat{\lambda}_{Jar}(\mathbf{X}; d) - \lambda_{Jar}(\mathbf{X}; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_{J\lambda r}(\mathbf{X}; d))$.

Asymptotic confidence intervals are readily found. Let $\widehat{C}_{Jn}(\mathbf{X}; d)$ be the nonempty closed and bounded interval of real numbers x such that $\widehat{G}_{Jn}(x, \mathbf{X}; d) = \widehat{G}_{Jn-}(\mathbf{X}; d)$. Let $\widehat{c}_{Jn}(\mathbf{X}; d)$ be the midpoint of $\widehat{C}_{Jn}(\mathbf{X}; d)$. Let $s_{na}^2\left(\widehat{G}_{Jn-}(\mathbf{X}; d)\right)$ be n^{-2} times the sum of $\left[\eta_J\left(\widehat{c}_{Jn}(\mathbf{X}; d), \mathbf{X}_i; d\right) - \widehat{G}_{Jn-}(\mathbf{X}; d)\right]^2$ for $1 \leq i \leq n$. As n approaches ∞, $ns_{na}^2\left(\widehat{G}_{Jn}(\mathbf{X}; d)\right)$ converges to $\sigma^2(\eta_J(c_J(\mathbf{X}; d), \mathbf{X}; d))$ with probability 1. If $\eta_J(c_J(\mathbf{X}; d), \mathbf{X}; d)$ does not equal $G_{J-}(\mathbf{X}; d)$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$|\, G_{J-}(\mathbf{X}; d) - \widehat{G}_{Jn-}(\mathbf{X}; d)\, | \leq z_{\alpha/2} s_{na}\left(\widehat{G}_{Jn-}(\mathbf{X}; d)\right).$$

Let

$$\widehat{\eta}_{J\lambda n}\left(\mathbf{X}_i, \mathbf{X}; d\right) = \left[\widehat{G}_{Jn-}(\mathbf{X}; d)\right]^{-1}\left[d_J\left(\mathbf{X}_i\right) - \widehat{\lambda}_{Jan}(\mathbf{X}; d)\, \eta_J\left(\widehat{\mathbf{c}}_{Jn}(\mathbf{X}; \mathbf{d}), \mathbf{X}_i; d\right)\right]$$

for $1 \leq i \leq n$. Let $s_{na}^2\left(\widehat{\lambda}_{Jan}(\mathbf{X}; d)\right)$ be n^{-2} times the sum of $\left[\widehat{\eta}_{J\lambda n}\left(\mathbf{X}_i, \mathbf{X}; d\right)\right]^2$ for $1 \leq i \leq n$. With probability 1, as n approaches ∞, $ns_{na}^2\left(\widehat{G}_{Jn-}(\mathbf{X}; d)\right)$ converges to $\sigma^2(\eta_{J\lambda}(c_J(\mathbf{X}; d), \mathbf{X}; d))$. If $\eta_{J\lambda}(c_J(\mathbf{X}; d), \mathbf{X}; d))$ is not 0 with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$|\, \lambda_{Ja}(\mathbf{X}; d) - \widehat{\lambda}_{Jan}(\mathbf{X}; d)\, | \leq z_{\alpha/2} s_{na}\left(\widehat{\lambda}_{Jan}(\mathbf{X}; d)\right).$$

In the nominal case, let $\widehat{C}_{Jrn}(\mathbf{X}; d)$ be the nonempty closed and bounded convex set of $\mathbf{p}$ in Π_r such that $\widehat{G}_{Jrn}\left(\mathbf{p}, \mathbf{X}; d\right) = \widehat{G}_{Jrn-}(\mathbf{X}; d)$. Let $\widehat{c}_{Jrn}(\mathbf{X}; d)$ in $\widehat{C}_{Jrn}(\mathbf{X}; d)$ be a random vector. Let $s_{na}^2\left(\widehat{G}_{Jrn-}(\mathbf{X}; d)\right)$ be n^{-2} times the sum of $\left[\eta_{Jr}\left(\widehat{c}_{Jrn}(\mathbf{X}; d), \mathbf{X}_i; d\right) - \widehat{G}_{Jrn-}(\mathbf{X}; d)\right]^2$ for $1 \leq i \leq n$. As n approaches ∞, $ns_{na}^2\left(\widehat{G}_{Jrn}(\mathbf{X}; d)\right)$ converges to $\sigma^2(\eta_{Jr}(\mathbf{c}_{Jr}(\mathbf{X}; d))$ with probability 1. If $\eta_{Jr}(\mathbf{c}_{Jr}(\mathbf{X}; d))$ does not equal $G_{Jr-}(\mathbf{X}; d)$ with probability 1, then the probability approaches $1 - \alpha$ that

$$|\, G_{Jr-}(\mathbf{X}; d) - \widehat{G}_{Jrn-}(\mathbf{X}; d)\, | \leq z_{\alpha/2} s_{na}\left(\widehat{G}_{Jrn-}(\mathbf{X}; d)\right).$$

Let

$$\widehat{\eta}_{J\lambda rn}\left(\mathbf{X}_i; d\right) = \left[\widehat{G}_{Jrn-}(\mathbf{X}; d)\right]^{-1}\left[d_{Jr}\left(\mathbf{X}_i\right) - \widehat{\lambda}_{Jarn}(\mathbf{X}; d)\, \eta_{J\lambda r}\left(\widehat{c}_{Jrn}(\mathbf{X}; d), \mathbf{X}_i; d\right)\right]$$

for $1 \le i \le n$. Let $s^2_{na}\left(\hat{\lambda}_{Jarn}(\mathbf{X}; d)\right)$ be n^{-2} times the sum of $\left[\hat{\eta}_{J\lambda rn}\left(\mathbf{X}_i; d\right)\right]^2$ for $1 \le i \le n$. As n approaches ∞, $ns^2_{na}\left(\hat{G}_{Jrn}(\mathbf{X}; d)\right)$ converges to $\sigma^2(\eta_{Jr}(\mathbf{c}_{Jr}(\mathbf{X}; d))$ with probability 1. If $\eta_{J\lambda r}(\mathbf{c}_{Jr}(\mathbf{X}; d))$ does not equal 0 with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$\mid \lambda_{Jar}(\mathbf{X}; d) - \hat{\lambda}_{Jarn}(\mathbf{X}; d) \mid \le z_{\alpha/2} s_{na}\left(\hat{\lambda}_{Jarn}(\mathbf{X}; d)\right).$$

Measures of Prediction Accuracy

Results for estimation of measures of prediction depend on the set of functions permitted as predictors and on the distribution of Z. In this problem, mutually independent pairs $(Y_i, \mathbf{Z}_i)$, $i \ge 1$, are considered such that each $(Y_i, \mathbf{Z}_i)$ has the same distribution as $(Y, \mathbf{Z})$. It is assumed that Y is in $L_{h(d)}$ and no real y exists such that $Y = y$ with probability 1. It is also assumed that no closed and convex set with empty interior includes $\mathbf{Z}$ with probability 1. Let $\hat{G}_n(Y \mid \mathbf{Z}; d)$ be the nonnegative function on $L_{h(d)}(\mathbf{Z})$ with value $\hat{G}_n\left(f, Y \mid \mathbf{Z}; d\right) = n^{-1}\sum_{i=1}^{n} d\left(Y_i - f\left(\mathbf{Z}_\mathrm{i}\right)\right)$ at f in $L_{h(d)}(\mathbf{Z})$, Let $\hat{G}_{n-}(Y \mid \mathbf{Z}; F, d) \ge 0$ be the infimum of $\hat{G}_n(Y \mid \mathbf{Z}; d)$ on F. Let $\hat{G}_n(Y; d)$ be the real function on the real line with value $\hat{G}_n\left(y, Y; d\right) = n^{-1}\sum_{i=1}^{n} d\left(Y_i - y\right)$ at y real, and let $\hat{G}_{n-}(Y; d)$ be the infimum of $\hat{G}_n(Y; d)$. Let $\hat{\lambda}_n(Y \mid \mathbf{Z}; F, d) \ge 0$ be $\left[\hat{G}_{n-}(Y; d) - \hat{G}_{n-}(Y \mid \mathbf{Z}; F, d)\right] / \hat{G}_{n-}(Y; d)$.

Several cases are straightforward (Haberman, 1989). It is always the case that, as n approaches ∞, $\hat{G}_{n-}(Y; d)$ converges with probability 1 to $G-(Y; d)$. For asymptotic normality, if a unique $c(Y; d)$ exists such that $G(c(Y; d), Y; d) = G-(Y; d)$ and Y is in $L_{2h(d)}$, then, as n approaches ∞, $n^{1/2}\left[\hat{G}_{n-}(Y; d) - G_{-}(Y; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(d(Y - c(Y; d)))$. Consider the nonempty closed bounded interval $\hat{C}_n(Y; d)$ of real c such that $\hat{G}_n(c, Y; d) = \hat{G}_{n-}(Y; d)$. Let $\hat{c}_n(Y; d)$ be the midpoint of $\hat{C}_n(Y; d)$. Let $s^2_{na}\left(\hat{G}_{n-}(Y; d)\right)$ be n^{-2} times the sum of $\left[d\left(Y_i - \hat{c}_n(Y; d)\right) - \hat{G}_{n-}(Y; d)\right]^2$ for $1 \le i \le n$. As n approaches ∞, $ns^2_{na}\left(\hat{G}_{n-}(Y; d)\right)$ converges to $\sigma^2(d(Y - c(Y; d)))$ with probability 1. If Y is not equal to $c(Y; d)$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$\mid G_{-}(Y; d) - \hat{G}_{n-}(Y; d) \mid \le z_{\alpha/2} s_{na}\left(\hat{G}_{n-}(Y; d)\right).$$

If $F = L_{h(d)}(\mathbf{Z})$, $\mathcal{Z}$ is finite, and $P(\mathbf{Z} = \mathbf{z}) > 0$ for all $\mathbf{z}$ in $\mathcal{Z}$, then, as n approaches ∞, $\hat{G}_{n-}(Y \mid \mathbf{Z}; d) = \hat{G}_{n-}\left(Y|\mathbf{Z}; L_{h(d)}(\mathbf{Z}), d\right)$ converges to $G_{-}(Y \mid \mathbf{Z}; d)$ with probability 1. In addition, $\hat{\lambda}_n(Y \mid \mathbf{Z}; d) = \hat{\lambda}_n\left(Y|\mathbf{Z}; L_{h(d)}(\mathbf{Z}), d\right)$ converges to $\lambda(Y \mid \mathbf{Z}; d)$ with probability 1. In the case of normal approximations, let Y be in $L_{2h(d)}$ and let $c(Y; d)$ be the only member of $C(Y; d)$. For each $\mathbf{z}$ in $\mathcal{Z}$, let a unique $c(Y \mid \mathbf{Z} = \mathbf{z}; d)$ exist such that $G(c(Y \mid \mathbf{Z} = \mathbf{z}; d), Y \mid \mathbf{Z} = \mathbf{z}; d) = G_{-}(Y \mid \mathbf{Z} = \mathbf{z}; d)$. Let $c(Y \| \mathbf{Z}; d)$ be $c(Y \mid \mathbf{Z} = \mathbf{z}; d)$ if $\mathbf{Z} = \mathbf{z}$ in $\mathcal{Z}$. As n approaches ∞, $n^{1/2}\left[\hat{G}_{n-}(Y \mid \mathbf{Z}; d) - G_{-}(Y \mid \mathbf{Z}; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(d(Y - c(Y \| \mathbf{Z}; d)))$.

For an asymptotic confidence interval, let $\mathcal{Z}_n$ be the set of $\mathbf{z}$ in $\mathcal{Z}$ such that, for some positive integer $i \le n$, $\mathbf{Z}_i = \mathbf{z}$. For $\mathbf{z}$ in $\mathcal{Z}_n$, let $\hat{C}_n(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the nonempty, bounded, and closed interval of real c such $\hat{G}_n(c, Y \mid \mathbf{Z} = \mathbf{z}; d) = \hat{G}_{n-}(Y \mid \mathbf{Z} = \mathbf{z}; d)$, where $\hat{G}_{n-}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ is the infimum of the real function $\hat{G}_n(Y \mid \mathbf{Z} = \mathbf{z}; d)$ with value for real y of $\hat{G}_n\left(y, Y \mid \mathbf{Z} = \mathbf{z}; d\right)$ equal to the average of $d(Y_i - y)$ for positive integers $i \le n$ such that $\mathbf{Z}_i = \mathbf{z}$. Let $\hat{c}_n(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the midpoint of $\hat{C}_n(Y \mid \mathbf{Z} = \mathbf{z}; d)$. Let $s^2_{na}\left(\hat{G}_{n-}(Y; d)\right)$ be n^{-2} times the sum of $\left[d\left(Y_i - \hat{c}_n\left(Y|\mathbf{Z} = \mathbf{Z}_i; d\right)\right) - \hat{G}_{n-}\left(Y|\mathbf{Z} = \mathbf{Z}_i; d\right)\right]^2$ for $1 \le i \le n$. As n approaches ∞, $ns^2_{na}\left(\hat{G}_{n-}(Y \mid \mathbf{Z}; d)\right)$ converges to $\sigma^2(d(Y - c(Y \mid \mathbf{Z}; d)))$ with probability 1. If Y is not equal to $c(Y \| \mathbf{Z}; d)$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$\mid G_{-}(Y \mid \mathbf{Z}; d) - \hat{G}_{n-}(Y \mid \mathbf{Z}; d) \mid \le z_{\alpha/2} s_{na}\left(\hat{G}_{n-}(Y \mid \mathbf{Z}; d)\right).$$

In the case of $\lambda(Y \mid \mathbf{Z}; d)$, let

$$\eta_{\lambda}(Y \mid \mathbf{Z}; d) = \left[G_{-}(Y; d)\right]^{-1}\left[d(Y - c(Y; d)) - \lambda(Y \mid \mathbf{Z}; d)\, d(Y - c(Y \| \mathbf{Z}; d))\right].$$

As n approaches ∞, $n^{1/2}\left[\hat{\lambda}_n(Y \mid \mathbf{Z}; d) - \lambda(Y \mid \mathbf{Z}; d)\right]$ converges in law to a normal random variable with mean 0 and variance $\sigma^2(\eta_{\lambda}(Y \mid \mathbf{Z}; d))$.

If

$$\hat{\eta}_{\lambda in}(Y \mid \mathbf{Z}; d) = \left[\hat{G}_{n-}(Y; d)\right]^{-1}\left[d\left(Y_i - \hat{c}_n(Y; d)\right) - \hat{\lambda}_n(Y \mid \mathbf{Z}; d)\, d\left(Y_i - \hat{c}_n\left(Y|\mathbf{Z} = \mathbf{Z}_i; d\right)\right)\right]$$

and $s^2_{na}\left(\hat{\lambda}_n(Y \mid \mathbf{Z}; d)\right)$ is n^{-2} times the sum of $\left[\hat{\eta}_{\lambda in}(Y \mid \mathbf{Z}; d)\right]^2$ for $1 \le i \le n$, then, as n approaches ∞, $ns^2_{na}\left(\hat{\lambda}_n(Y \mid \mathbf{Z}; d)\right)$ converges to $\sigma^2(\eta(Y \mid \mathbf{Z}; d))$ with probability 1. If $d(Y - c(Y))$ does not equal $\lambda(Y \mid \mathbf{Z}; d)d(Y - c(Y\|\mathbf{Z}; d))$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$\mid \lambda(Y \mid \mathbf{Z}; d) - \hat{\lambda}_n(Y \mid \mathbf{Z}; d) \mid \le z_{\alpha/2} s_{na}\left(\hat{\lambda}_n(Y \mid \mathbf{Z}; d)\right).$$

The case of $F = A(\mathbf{Z})$ for Z_k in $L_{h(d)}$ for $1 \le k \le K$ is relatively straightforward, even without any assumption that $\mathcal{Z}$ is finite. As n approaches ∞, $\hat{G}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d)$ converges to $G(Y|\mathbf{Z}; A(\mathbf{Z}), d)$ with probability 1, and $\hat{\lambda}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d)$ converges to $\lambda(Y|\mathbf{X}; A(\mathbf{Z}), d)$ with probability 1. If a unique f in $A(\mathbf{Z})$ exists such that $G(f, Y \mid \mathbf{Z}; d) = G_-(Y|\mathbf{Z}; A(\mathbf{Z}), d)$ and Y and Z_k, $1 \le k \le K$, are in $L_{2h(d)}$, then a normal approximation is available. As n approaches ∞, $n^{1/2}\left[\hat{G}_{n-}(Y|\mathbf{Z}; A(\mathbf{Z}), d) - G(Y|\mathbf{Z}; A(\mathbf{Z}); d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(d(Y - f(\mathbf{Z})))$. In addition, if $c(Y; d)$ is the only element of $C(Y; d)$, then, as n approaches ∞, $n^{1/2}\left[\hat{\lambda}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d) - \lambda(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_\lambda(Y|\mathbf{Z}; A(\mathbf{Z}), d))$, where

$$\eta_\lambda(Y|\mathbf{Z}; A(\mathbf{Z}), d) = \left[G_-(Y; d)\right]^{-1}\left[d(Y - c(Y; d)) - \lambda(Y|\mathbf{Z}; A(\mathbf{Z}); d)\, d\left(Y - f(\mathbf{Z})\right)\right].$$

For an asymptotic confidence interval, let $\hat{F}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d)$ be the set of f in $A(\mathbf{Z})$ such that $\hat{G}_n\left(f, Y \mid \mathbf{Z}; d\right) = \hat{G}_{n-}\left(f, Y \mid \mathbf{Z}; A(\mathbf{Z}), d\right)$. Let $\hat{f}_n$ in $\hat{F}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d))$ be defined so that $\hat{f}_n(\mathbf{Z}_i)$ is a random vector for $1 \le i \le n$. Let $s^2_{na}\left(\hat{G}_{n-}(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right)$ be n^{-2} times the sum of $\left[d\left(Y - \hat{f}_n(\mathbf{Z}_i)\right) - \hat{G}_{n-}(Y \mid \mathbf{Z}; d)\right]^2$ for $1 \le i \le n$. As n approaches ∞, $ns^2_{na}\left(\hat{G}_{n-}(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right)$ converges to $\sigma^2(d(Y - f(\mathbf{Z})))$ with probability 1. If $d(Y - f(\mathbf{Z}))$ does not equal $G_-(Y \mid \mathbf{Z}; d)$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$\mid G_-(Y|\mathbf{Z}; A(\mathbf{Z}), d) - \hat{G}_{n-}(Y|\mathbf{Z}; A(\mathbf{Z}), d) \mid \le z_{\alpha/2} s_{na}\left(\hat{G}_{n-}(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right).$$

If

$$\hat{\eta}_{\lambda in}(Y|\mathbf{Z}; A(\mathbf{Z}), d) = \left[\hat{G}_{n-}(Y; d)\right]^{-1}\left[d\left(Y_i - \hat{c}_n(Y; d)\right) - \hat{\lambda}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d)\, d\left(Y_i - \hat{f}_n(\mathbf{Z}_i)\right)\right)]$$

and $s^2_{na}\left(\hat{\lambda}_n(Y \mid \mathbf{Z}; d)\right)$ is n^{-2} times the sum of $\left[\hat{\eta}_{\lambda in}(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right]^2$ for $1 \le i \le n$, then, as n approaches ∞, $ns^2_{na}\left(\hat{\lambda}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right)$ converges to $\sigma^2(\eta_\lambda(Y|\mathbf{Z}; A(\mathbf{Z}), d))$ with probability 1. If $\eta_\lambda(Y \mid \mathbf{Z}; A(\mathbf{Z}; d)$ is not 0 with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$\mid \lambda(Y|\mathbf{Z}; A(\mathbf{Z}), d) - \hat{\lambda}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d) \mid \le z_{\alpha/2} s_{na}\left(\hat{\lambda}_n(Y|\mathbf{Z}; A(\mathbf{Z}), d)\right).$$

For the case of probability prediction, let Y always be in $\bar{r}$. Let $\hat{G}_{rn}(Y \mid \mathbf{Z}; d)$ be the nonnegative function on $B_r(\mathbf{Z})$ with value $\hat{G}_{rn}(\mathbf{f}, Y \mid \mathbf{Z}; d) = n^{-1}\sum_{i=1}^n \eta_r\left(\mathbf{f}\left(\mathbf{Z_i}\right), Y_i; d\right)$ at $\mathbf{f}$ in $B_r(\mathbf{Z})$. Let $\hat{G}_{rn-}\left(Y|\mathbf{Z}; F_r, d\right) \ge 0$ be the infimum of $\hat{G}_{nr}(Y \mid \mathbf{Z}; d)$ on F_r. Let $\hat{G}_{rn}(Y; d)$ be the real function on Π_r with value $\hat{G}_{rn}\left(y, Y; d\right) = n^{-1}\sum_{i=1}^n \eta_r\left(\mathbf{p}, Y_i; d\right)$ at $\mathbf{p}$ in Π_r, and let $\hat{G}_{rn-}(Y; d)$ be the infimum of $\hat{G}_{rn}(Y; d)$. Let $\hat{\lambda}_{rn}\left(Y|\mathbf{Z}; F_r, d\right) \ge 0$ be $\left[\hat{G}_{rn-}(Y; d) - \hat{G}_{rn-}\left(Y|\mathbf{Z}; F_r, d\right)\right] / \hat{G}_{rn-}(Y; d)$. Let $\hat{G}_{rn-}(Y \mid \mathbf{Z}; d) = \hat{G}_{rn-}\left(Y|\mathbf{Z}; B_r(\mathbf{Z}), d\right)$ and $\hat{\lambda}_{rn-}(Y \mid \mathbf{Z}; d) = \hat{\lambda}_{rn-}\left(Y|\mathbf{Z}; B_r(\mathbf{Z}), d\right)$.

It is always the case that $\hat{G}_{rn-}(Y; d)$ converges with probability 1 to $G_{r-}(Y; d)$. If $C_r(Y; d)$ contains one element $\mathbf{p}_r(Y; d)$, then, as n approaches ∞, $n^{1/2}\left[\hat{G}_{rn-}(Y; d) - G_{r-}(Y; d)\right]$ converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_r(\mathbf{p}_r(Y; d), Y; d))$.

An asymptotic confidence interval is also available. Consider the nonempty closed convex set $\hat{P}_{rn}(Y; d)$ of $\mathbf{p}$ in Π_r such that $\hat{G}_{rn}\left(\mathbf{p}, Y; d\right) = \hat{G}_{rn-}(Y; d)$. Let $\hat{\mathbf{p}}_n(Y; d)$ be a random vector with values in $\hat{P}_{rn}(Y; d)$. Let $s^2_{na}\left(\hat{G}_{rn-}(Y; d)\right)$ be n^{-2} times the sum of for $1 \le i \le n$. Then $ns^2_{na}\left(\hat{G}_{rn-}(Y; d)\right)$ converges to $\sigma^2(\eta_r(\mathbf{p}_r(Y; d), Y; d))$ with probability 1.

If $\eta_r(\mathbf{p}_r(Y; d), Y; d)$ is not equal to $G_{r-}(Y; d)$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$|G_{r-}(Y; d) - \widehat{G}_{rn-}(Y; d)| \leq z_{\alpha/2} s_{na}\left(\widehat{G}_{rn-}(Y; d)\right).$$

If $\mathcal{Z}$ is finite and $\mathbf{Z} = \mathbf{z}$ with positive probability for each $\mathbf{z}$ in $\mathcal{Z}$, then $\widehat{G}_{rn-}(Y \mid \mathbf{Z}; d)$ converges to $G_{r-}(Y \mid \mathbf{Z}; d)$ with probability 1 and $\widehat{\lambda}_{rn-}(Y \mid \mathbf{Z}; d)$ converges to $\lambda_r(Y \mid \mathbf{Z}; d)$ with probability 1 as n approaches 1. For a normal approximation, let $\mathbf{p}_r(Y; d)$ be the only member of $C_r(Y; d)$, and, for $\mathbf{z}$ in $\mathcal{Z}$, let $\mathbf{p}_r(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the unique member of $C_r(Y \mid \mathbf{Z} = \mathbf{z}; d)$. Let $\mathbf{p}_r(Y \| \mathbf{Z}; d)$ be $\mathbf{p}_r(Y \mid \mathbf{Z} = \mathbf{z}; d)$ if $\mathbf{Z} = \mathbf{z}$. As n approaches ∞,

$$n^{1/2}\left[\widehat{G}_{rn-}(Y \mid \mathbf{Z}; d) - G_r(Y \mid \mathbf{Z}; d)\right]$$

converges in law to a normal distribution with mean 0 and variance $\sigma^2(\eta_r(\mathbf{p}_r(Y \| \mathbf{Z}; d), Y; d))$. Let

$$\eta_{\lambda r}(Y \mid \mathbf{Z}; d) = \left[G_{r-}(Y; d)\right]^{-1}\left[\eta_r\left(\mathbf{p}_r(Y; d), Y; d\right) - \lambda_r(Y \mid \mathbf{Z}; d)\,\eta_r\left(\mathbf{p}_r(Y \mid \mathbf{Z}, Y; d)\right)\right].$$

Then $n^{1/2}\left[\widehat{\lambda}_{rn}(Y \mid \mathbf{Z}; d) - \lambda_r(Y \mid \mathbf{Z}; d)\right]$ converges in law to a normal random variable with mean 0 and variance $\sigma^2(\eta_{\lambda r}(Y \mid \mathbf{Z}; d))$.

For an asymptotic confidence interval, for $\mathbf{z}$ in $\mathcal{Z}_n$, let $\widehat{P}_{rn}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be the nonempty, bounded, and closed convex set of $\mathbf{p}$ in Π_r such $\widehat{G}_{rn}\left(\mathbf{p}, Y \mid \mathbf{Z} = \mathbf{z}; d\right) = \widehat{G}_{rn-}(Y \mid \mathbf{Z} = \mathbf{z}; d)$, where $\widehat{G}_{rn-}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ is the infimum of the real function $\widehat{G}_{rn}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ with value for $\mathbf{p}$ in Π_r of $\widehat{G}_{rn}\left(\mathbf{p}, Y \mid \mathbf{Z} = \mathbf{z}; d\right)$ equal to the average of $\eta_r(\mathbf{p}, Y_i; d)$ for positive integers $i \leq n$ such that $\mathbf{Z}_i = \mathbf{z}$. Let $\widehat{\mathbf{p}}_{rn}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ in $\widehat{P}_{rn}(Y \mid \mathbf{Z} = \mathbf{z}; d)$ be defined so that $\widehat{p}_{rn}\left(Y | \mathbf{Z}_i; d\right)$ is a random vector for each positive integer $i \leq n$. Let $s_{rna}^2\left(\widehat{G}_{rn-}(Y; d)\right)$ be n^{-2} times the sum of $\left[\eta_r\left(\widehat{\mathbf{p}}_{rn}\left(Y | \mathbf{Z}_i; d\right), Y_i; d\right) - \widehat{G}_{rn-}\left(Y | \mathbf{Z} = \mathbf{Z}_i; d\right)\right]^2$ for $1 \leq i \leq n$. As n approaches ∞, $ns_{na}^2\left(\widehat{G}_{rn-}(Y \mid \mathbf{Z}; d)\right)$ converges to $\sigma^2(\eta_r(\mathbf{p}_r(Y \| \mathbf{Z}; d)))$ with probability 1. If Y is not equal to $c(Y \| \mathbf{Z}; d)$ with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$|G_-(Y \mid \mathbf{Z}; d) - \widehat{G}_{n-}(Y \mid \mathbf{Z}; d)| \leq z_{\alpha/2} s_{na}\left(\widehat{G}_{n-}(Y \mid \mathbf{Z}; d)\right).$$

If

$$\widehat{\eta}_{\lambda rin}(Y \mid \mathbf{Z}; d) = \left[\widehat{G}_{rn-}(Y; d)\right]^{-1}\left[\eta_r\left(\widehat{\mathbf{p}}_{rn}(Y; d) - \widehat{\lambda}_{rn}(Y \mid \mathbf{Z}; d)\,\eta_r\left(\widehat{\mathbf{p}}_{rn}\left(\mathbf{Z} = \mathbf{Z}_i; d\right), Y; d\right)\right)\right]$$

and $s_{na}^2\left(\widehat{\lambda}_{rn}(Y \mid \mathbf{Z}; d)\right)$ is n^{-2} times the sum of $\left[\widehat{\eta}_{\lambda rin}(Y \mid \mathbf{Z}; d)\right]^2$ for $1 \leq i \leq n$, then, as n approaches ∞, $ns_{na}^2\left(\widehat{\lambda}_{rn}(Y \mid \mathbf{Z}; d)\right)$ converges to $\sigma^2(\eta_{\lambda r}(Y \mid \mathbf{Z}; d))$ with probability 1. If $\eta_{\lambda r}(Y \mid \mathbf{Z}; d)$ is not 0 with probability 1, then, as n approaches ∞, the probability approaches $1 - \alpha$ that

$$|\lambda_r(Y \mid \mathbf{Z}; d) - \widehat{\lambda}_n(Y \mid \mathbf{Z}; d)| \leq z_{\alpha/2} s_{na}\left(\widehat{\lambda}_{rn}(Y \mid \mathbf{Z}; d)\right).$$

The case of $F_r = B_r(\mathbf{h}, \mathbf{Z})$ based on the H-dimensional function $\mathbf{h}$ on $\mathcal{Z}$ such that $\mathbf{h}(\mathbf{Z})$ is a random vector involves no special considerations. Here $\widehat{G}_{rn-}\left(Y \mid \mathbf{Z}; B_r(\mathbf{h}, \mathbf{Z}), d\right) = \widehat{G}_{rn-}(Y | \mathbf{h}(\mathbf{Z}); d)$ and $\widehat{\lambda}_{rn}\left(Y \mid \mathbf{Z}; B_r(\mathbf{h}, \mathbf{Z}), d\right) = \widehat{\lambda}_{rn}(Y | \mathbf{h}(\mathbf{Z}), d)$.

Concluding Remarks

Measures discussed in this report apply both to discrete and continuous variables, except that the case of nominal predicted variables requires supplementary treatment. This observation appears not to be widely recognized for measures of agreement, so that many statistical packages only treat measures of agreement for random variables with integer values. Given that quadratic discrepancy generally has more attractive properties than alternatives and is tied to the familiar concepts of variance, correlation, and the coefficient of determination, it appears that this measure should be given preference without compelling reason to act differently.

Without strong reasons to act otherwise, proportional reduction in error should receive emphasis in reporting. Although kappa measures are traditional and do provide proportional reduction in error, it is fair to ask whether lambda measures of agreement provide a better understanding of poor agreement.

Although results have been provided for agreement and prediction of nominal variables, it is fair to consider how often these measures are really appropriate when the variable assumes more than two values. The issue is how often values of a nominal variable can be regarded as equally distant from each other.

One further matter should be noted in the case of probability prediction. The approach here emphasizes discrepancy measures d that are suitable both for measures of agreement and for measures of prediction. For probability prediction, there is a long tradition of using the log penalty function in which a penalty of $-\log p_Y$ is recorded if Y occurs and the predicted probability vector $\mathbf{p}$ in Π_r has elements p_y for y in $\bar{r}$. This option has an extensive literature of its own (Goodman & Kruskal, 1959; Haberman, 1982a, 1989), among numerous other references. The log penalty is closely linked to maximum likelihood, so estimation is generally quite efficient, especially for families of functions based on multinomial logits (Haberman, 1982a).

References

Berk, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Annals of Mathematical Statistics*, *43*, 193–204. https://doi.org/10.1214/aoms/1177692713

Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Annals of Mathematical Statistics*, *18*, 105–110. https://doi.org/10.1214/aoms/1177730497

Bloomfield, P., & Steiger, W. (1983). *Least absolute deviations: Theory, applications and algorithms*. Boston, MA: Birkhäuser.

Brown, L. D., & Purves, R. (1973). Measurable selections of Extrema. *Annals of Statistics*, *1*, 902–912. https://doi.org/10.1214/aos/1176342510

Cicchetti, D. V., & Allison, T. (1971). A new procedure for assessing reliability of scoring EEG sleep recordings. *The American Journal of EEG Technology*, *11*, 101–110. https://doi.org/10.1080/00029238.1971.11080840

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220. https://doi.org/10.1037/h0026256

Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin*, *88*, 322–328. https://doi.org/10.1037/0033-2909.88.2.322

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review*, *30*, 341–353. https://doi.org/10.2307/2090715

Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.). Edinburgh, Scotland: Oliver and Boyd.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613–619. https://doi.org/10.1177/001316447303300309

Fleiss, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin*, *72*, 323–327. https://doi.org/10.1037/h0028106

Gilula, Z., & Haberman, S. J. (1995a). Dispersion of categorical variables and penalty functions: Derivation, estimation, and comparability. *Journal of the American Statistical Association*, *90*, 1447–1452. https://doi.org/10.2307/2291537

Gilula, Z., & Haberman, S. J. (1995b). Prediction functions for categorical panel data. *Annals of Statistics*, *23*, 1130–1142. https://doi.org/10.1214/aos/1176324701

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–764. https://doi.org/10.2307/2281536

Goodman, L. A., & Kruskal, W. H. (1959). Measures of association for cross classifications. II: Further discussion and references. *Journal of the American Statistical Association*, *54*, 123–163. https://doi.org/10.2307/2282143

Goodman, L. A., & Kruskal, W. H. (1963). Measures of association for cross classifications III: Approximate sampling theory. *Journal of the American Statistical Association*, *58*, 310–364. https://doi.org/10.2307/2283271

Goodman, L. A., & Kruskal, W. H. (1972). Measures of association for cross classifications, IV: Simplification of asymptotic variances. *Journal of the American Statistical Association*, *67*, 415–421. https://doi.org/10.2307/2284396

Haberman, S. J. (1982a). Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association*, *77*, 568–580. https://doi.org/10.2307/2287713.

Haberman, S. J. (1982b). Measures of association. In S. Kotz & N. L. Johnson (Eds.)., *Encyclopedia of statistical sciences* (Vol. *1*, pp. 130–137). New York, NY: John Wiley. https://doi.org/10.1002/0471667196.ess0066

Haberman, S. J. (1989). Concavity and estimation. *Annals of Statistics*, *17*, 1631–1661. https://doi.org/10.1214/aos/1176347385

Haberman, S. J. (1996). *Advanced statistics*. New York, NY: Springer. https://doi.org/10.1007/978-1-4757-4417-0

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, *19*, 293–325. https://doi.org/10.1214/aoms/1177730196

Hoeffding, W. (1961). *The strong law of large numbers for U-statistics* (Tech. Rep. No. 302). Chapel Hill, NC: Institute of Statistics, University of North Carolina.

Lee, A. J. (1990). *U-statistics: Theory and practice*. New York, NY: Marcel Dekker.

Lehmann, E. L. (1986). *Testing statistical hypotheses* (2nd ed.). New York, NY: Springer. https://doi.org/10.1007/978-1-4757-1923-9

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Menlo Park, NJ: Addison-Wesley.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400873173

Wright, S. (1920). Correlation and causation. *Journal of Agricultural Research*, *20*, 557–585. https://doi.org/10.1073/pnas.6.6.320

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, *60*, 812–854. https://doi.org/10.2307/2979746

Suggested citation:

Haberman, S. J. (2019). *Measures of agreement versus measures of prediction accuracy* (Research Report No. RR-19-20). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12258

Action Editor: Rebecca Zwick

Reviewers: Matthew Johnson